AD_____

Award Number: DAMD17-02-1-0346

TITLE: Scanning the Human Genome for Novel Therapeutic Targets for Breast Cancer

PRINCIPAL INVESTIGATOR: Gregory Hannon, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

REPORT DATE: April 2004

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2004	Annual (1 Apr 03-31 Mar 04)

4. TITLE AND SUBTITLE
Scanning the Human Genome for Novel Therapeutic Targets for Breast Cancer

5. FUNDING NUMBERS
DAMD17-02-1-0346

6. AUTHOR(S)
Gregory Hannon, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

E-Mail: hannon@cshl.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates. All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

The broad goal of this project is to develop genome-wide RNAi approaches in mammals and to apply these to the discovery of new therapeutic targets for cancer. Specifically, we will generate a library of short hairpin RNA expression constructs (shRNA) that ultimately correspond to every gene in the human genome. These will be made available as a public resource and used internally to screen for genes that are essential to the survival of breast cancer cells but which are dispensable for the survival of normal cells. A subset of these might prove suitable as therapeutic targets for breast cancer therapy. During the first year of funding, two things have become clear. First, although they were not in place at the time of submitting this application, we have largely developed the technologies necessary to pursue the above goal. Second, funding in the Innovator award falls far short of that necessary to achieve the goal. Relevant to the last point, we have been able to leverage the Innovator award with several other funding sources to crate a program, which is capable of meeting the proposed goal.

14. SUBJECT TERMS
Cancer Biology, Genetics, Synthetic Lethality, Apoptosis

15. NUMBER OF PAGES
60

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

The broad goal of this project is to develop genome-wide RNAi approaches in mammals and to apply these to the discovery of new therapeutic targets for cancer. Specifically, we have generated and continue to build a library of short hairpin RNA expression constructs (shRNA) that ultimately correspond to every gene in the human and mouse genomes. These will be are presently available as a public resource and used internally to screen for genes that are essential to the survival of breast cancer cells but which are dispensable for the survival of normal cells. A subset of these might prove suitable as therapeutic targets for breast cancer therapy. During the course of funding, two things have become clear. First, although they were not in place at the time of submitting this application, we have largely developed the technologies necessary to pursue the above goal. Second, funding in the Innovator award falls far short of that necessary to achieve the goal. Relevant to the last point, we have been able to leverage the Innovator award with several other funding sources to create a program, which is capable of meeting the proposed goal.

BODY

Progress toward developing the technology necessary for genome-wide RNAi in mammals (these were funded in part by the Innovator award and also by P01 from the NCI)

Studies on the mechanism of RNAi

All of the technology described above was built upon studies of the RNAi mechanism. While these studies are funded by an R01, they benefit from the Innovator award, and it is acknowledged as general support for the P.I. We have made progress relevant to this goal, understanding in more depth the biochemistry of the RNAi pathway and working to apply this knowledge to the improvement of RNAi as a tool.

A genome-wide RNAi library

Over the last 18 months, there have emerged two major methods for triggering RNAi in mammalian cells. These are transient silencing using siRNAs or stable or transient silencing using shRNAs. Both of these approaches have been validated in numerous publications. In considering how to construct a genome-wide RNA library for human cells, we examined both options. Our choice of the latter reflects several factors. First and foremost, shRNA expression constructs can be propagated and thus provide a limitless supply of material for public distribution. Second, many phenotypes, especially those relevant to breast cancer, require examination of cells over a long time frame. Third, shRNAs offer the flexibility to examine the consequences of silencing both in vitro and in vivo. Generation of the library is proceeding as a phased project

with funding coming in part from the Innovator award and in part from other public and commercial sources (NCI, Merck, Oncogene Sciences, Genetech). No funding mechanism has been permitted to place any restrictions on library distribution.

Construction of a first-generation, genome-wide RNAi library

The major accomplishment of the last year has been the production of our first generation RNAi library, covering approximately 10,000 human genes with some 28,000 sequence-verified shRNAs. These are built in the vector described below. We tested a subset of the library for its ability to score in a screen for defects in proteasome-mediated degradation. From ~7,000 library clones, we found ~100 that could inhibit proteasome function. Notably, 22 of these were known proteasome components. This work has now been published in Nature and is attached as Appendix 1.

shRNA optimization

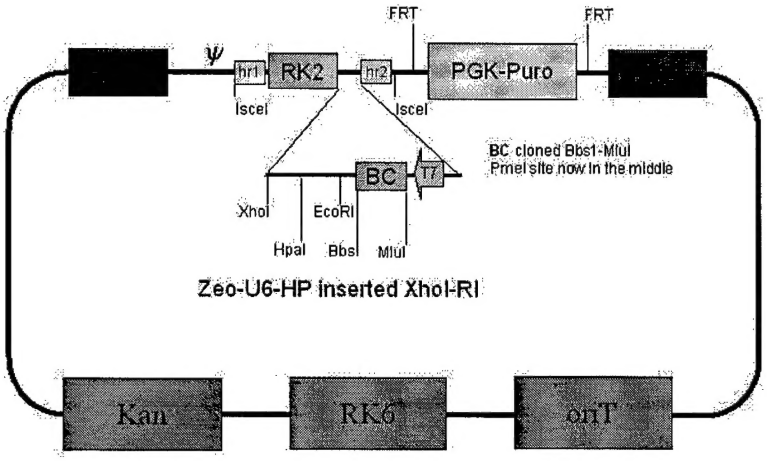
Our original series of vectors used a 29 base shRNA with a simple 4 base loop. Biochemical studies have revealed that natural miRNAs are processed into mature miRNAs through a two-step pathway that involves two specialized RNAseIII enzymes. We have found that by designing shRNAs as substrates for both components of the pathway that we can increase the amount of shRNA produced by our vectors in vivo by more than 10-fold. Additionally, detailed biochemical analysis of Dicer cleavage (see Appendix 2, submitted MS by Siolas et al) has allowed us to apply informatic strategies for picking more efficient shRNA sequences. These innovations have been incorporated into our second generation library (see below).

Vector construction and validation – vectors remain the same for both version 1 and version 2 libraries

In collaboration with Steve Elledge (Baylor), we constructed a flexible vector system for harboring the shRNA library. We have demonstrated that this vector can transfer shRNA inserts to a recipient plasmid by bacterial mating with ~100% efficiency. We have also validated transfers in multi-well formats suitable for moving subsets of or even the entire library. The original vector design had to be modified to remove loxP sites to avoid intellectual property restrictions placed upon us by Dupont. Persistant, although solvable problems, include the use of Zeocin and the inclusion of FRT sites. We are in the process of negotiating with Salk and Invitrogen to overcome these barriers to distribution. We are also undertaking a pilot project to replace Zeocin with Chloramphenical resistance by recombination. If this becomes necessary, we don't see this creating a significant delay in distribution. At present, Invitrogen has agreed to a 5% royalty

to permit distribution and we are now finalizing that agreement. The second-generation library will not contain zeocin resistance.

Figure 1. Vector system for library construction



For construction of the second generation library, we have modified pSM1 (Figure1) slightly to incoporate flanking sequences from a naturally occurring human microRNA. We have also slightly rearranged the elements within the vector to increase its stability. The second generation vector, pSM2, is pictured below in Figure 2.

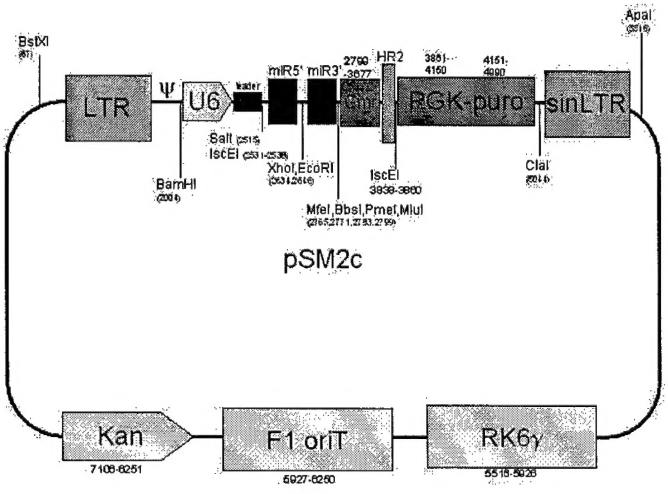


Figure 2. pSM2

shRNA library construction

Construction of the first generation library has been completed. In the end, 9,780 genes were covered with 28,000 different shRNAs. Given our progress in understanding the RNAi pathway and our application of this knowledge to the generation of better shRNA tools, we have undertaken the construction of second-generation libraries that will cover all known and predicted genes in human, mouse and rat. Thus far, we have sequence verified second generation constructs that cover ~16,000 human genes with ~23,000 shRNAs. The ability to carry out construction on such a large scale required the development of novel oligonucleotide synthesis strategies, which take advantage of highly parallel piezo-inket technology for oligo production. A manuscript describing this process is attached as Appendix 3. Unlike the first generation library, the public release of which awaited its publication, we are releasing the second generation library as it is completed. We have also constructed and are releasing about 5000 mouse clones, with the expansion of both mouse an human libraries occurring about ~10,000 shRNAs per month

Library Distribution

Our first and second generation libraries and vectors are now available to all academic investigators from Open Biosystems at very reasonable cost.

Library Screening

We are currently carrying out library screens for activation of several reporters. This aspect of the grant has been delayed somewhat by the substantial effort required to construct and sequence verify the some 60,000 clones that now comprise our shRNA resource.

KEY RESEARCH ACCOMPLISHMENTS

- Produced and publicly distributed a library of ~30,000 first-generation shRNAs
- Designed and released second-generation shRNA vectors
- Developed new oligonucleotide synthesis methods
- Validated rules-based approaches for shRNA production
- Began construction of second-generation libraries and released these to the public

REPORTABLE OUTCOMES

- Manuscripts (3 attached)
- Vector systems for RNAi (both transient and viral vectors, distributed to ~1000 labs so far)
- Additional funding for library construction from public and private sources
- First-generation human shRNA library (completed)
- Second-generation human library (in progress)

CONCLUSIONS

RNAi has emerged over the last two years as a powerful tool for experimental manipulation of gene expression and as a potential therapeutic strategy. We have made substantial progress toward validating the use of RNAi in mammals and have contributed key reagents to the scientific community.

APPENDIX 1

Synthetic shRNAs as highly potent RNAi triggers

Despina Siolas [1,2], Cara Lerner [3], Julja Burchard [3], Wei Ge[3], Patrick J. Paddison [2], Peter Linsley [3], Gregory J. Hannon [2*] and Michele A. Cleary [3*]

[1] Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[2] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY 11724

[3]Rosetta Inpharmatics LLC,
A Wholly Owned Subsidiary of Merck & Co., Inc.
401 Terry Ave North
Seattle, Washington 98109, USA

* to whom correspondence should be addressed
hannon@cshl.edu
Michele_cleary@merck.com

RNA interference continues to demonstrate its power as a transformational tool for mammalian genetics. Studies of the biochemical mechanisms underlying RNAi continue to hone our strategies for harnessing this process as an experimental tool. Here we show that chemically synthesized shRNAs specifically designed as Dicer substrates are more potent inducers of RNAi than siRNAs. Not only is maximal inhibition achieved at much lower levels of transfected RNA, but also endpoint inhibition is often greater. Mimicking natural pre-miRNAs by inclusion of a 2 nucleotide 3' overhang enhances the efficiency of Dicer cleavage and directs cleavage to a specific position in the precursor. Mapping of this processing site will permit the application of rules for siRNA design to shRNAs, both for chemical synthesis and vector-based delivery. Our data suggest an improved method for evoking RNAi in mammalian cells, with the hope that the ability to produce highly potent silencing triggers will ultimately ease the path toward therapeutic application of RNAi.

Many eukaryotic organisms respond to double-stranded RNA (dsRNA) by initiating a sequence-specific silencing pathway, known as RNA interference or RNAi. The ability to exploit RNAi as an experimental tool has evolved in lock-step with an elucidation of the underlying biochemical mechanism of this regulatory pathway. Harnessing RNAi for studies of gene function in invertebrate systems (e.g., *C. elegans* and *Drosophila*) grew out of the discovery that many potent effects of antisense RNA were actually triggered by the artefactual production of double-stranded RNA during *in vitro* transcription (*1*). In plants, use of RNAi as a genetic tool evolved following the realization that transgene co-suppression and virus-induced gene silencing (VIGS) relied on the generation of double-stranded RNA *in vivo* (*2*).

Due in part to the triggering of non-sequence specific responses by dsRNAs of greater than 30-50 nucleotides (reviewed in (*3*)), the experimental use of RNAi in mammalian systems awaited a much more detailed understanding of the RNAi mechanism. This emerged from analyses of plants that had mounted a co-suppression response, coupled with genetic studies in *C. elegans* and biochemical experiments in *Drosophila* cell and embryo extracts (reviewed in (*4*)). Initiation of RNAi occurs upon processing of double-stranded RNAs into ~22nt fragments, known as siRNAs (*5-7*), by an RNAseIII family nuclease, Dicer (*8*). These are then incorporated into an effector complex, RISC, which uses the small RNAs as a guide for selection and cleavage of complementary mRNAs (*6, 7, 9, 10*). The aforementioned findings led to the development of a methodology for experimentally programming the RNAi machinery in mammalian cells by direct delivery of chemically synthesized siRNAs. These duplexes of ~21 nt. consist of 19 paired bases with 2 nucleotide 3' overhangs and can be transfected directly into mammalian cells to produce a transient silencing response, the potency of which varies depending upon the siRNA sequence (*11-13*).

In many organisms, the RNAi machinery also serves an effector function for endogenous, non-coding RNAs, known as microRNAs (miRNAs) (reviewed in (*14*)). These are processed similarly to experimentally introduced dsRNAs, with two twists. First, miRNAs are initially generated as long primary transcripts (pri-miRNA), which are cleaved in the nucleus by another RNAseIII family member, Drosha (*15*). The liberated pre-miRNAs are exported to the cytoplasm where Dicer performs a second cleavage to produce small RNAs that are loaded into RISC (*16-19*). In the case of miRNAs, the cleavage sites are specific, and only a single, discrete sequence is liberated from the precursor (reviewed in (*14*)). These discoveries prompted the development of a second approach for triggering RNAi in mammalian cells using DNA vectors encoding short hairpin RNAs (shRNAs), modeled roughly after endogenous microRNAs (*20-25*). As with siRNAs, the efficacy of shRNAs varies with the specific sequence that is delivered.

Remarkably, for both miRNAs and siRNAs, the two strands of the processed dsRNA are treated unequally. Cloning efforts in a variety of

organisms yielded overwhelmingly one strand from most predicted miRNA precursors (*26-28*). A potential explanation for this outcome came from biochemical studies of siRNAs in *Drosophila* suggesting that relative thermodynamic instability at the 5' end of a given strand of the Dicer product favors its loading into RISC (*29*). This is in accord with analysis of predicted Dicer cleavage products of endogenous miRNAs (*30, 31*) and with studies of the efficacy of large numbers of siRNAs, which indicate greater suppression if the antisense strand (relative to the target mRNA) has an unstable 5' end (*31*). Recent reports have suggested that this loading might occur in a complex and be coordinated with Dicer cleavage (*32-34*). A possibility suggested by these mechanistic insights is that Dicer substrates might be more efficiently incorporated into RISC than siRNAs. We therefore sought to understand how Dicer processes shRNAs in order to permit comparison of the efficiency of silencing triggers that are predicted to produce equivalent RISC enzymes.

Dicer cleaves a single siRNA from the end of each shRNA

We began by producing ~70 chemically synthesized shRNAs, targeting various endogenous genes and reporters. We initially focused on a detailed analysis of one set of four shRNAs that target firefly luciferase (Fig 1a). The individual species differed in two distinct ways. First, the stems of the shRNAs were either 19 or 29 nucleotides in length. These sizes were chosen to reflect the two stem sizes most commonly used for expressed shRNAs *in vivo*. Second, each shRNA either contained or lacked a 2 nucleotide 3' overhang, identical to that produced by processing of pri-miRNAs by Drosha. Each species was end-labeled by enzymatic phosphorylation and incubated with recombinant human Dicer. The 29 nt. shRNA bearing the 3' overhang was converted almost quantitatively into a 22nt product by Dicer (Fig. 1b). In contrast, the 29 nt. shRNA that lacked the overhang generated very little 22 nt labeled product, although there was a substantial depletion of the starting material. Neither 19 nt. shRNA was cleaved to a significant extent by the Dicer enzyme. This result was not due to the lack of dsRNA in the 19nt shRNAs as all shRNA substrates were efficiently cleaved by bacterial RNAseIII (Fig 1c). Parallel analysis of identical shRNA substrates that were produced by in vitro transcription with T7 polymerase and uniformly labeled clarified the results obtained with end-labeled substrates (not shown). Specifically, 19 nt. shRNAs were not cleaved. However, both the overhung and the blunt 29 nucleotide shRNAs gave rise to 22 nt products, albeit at reduced levels in the latter case. These results suggest that Dicer requires a minimum stem length for productive cleavage. Furthermore, they are consistent with a hypothesis that the presence of a correct 3' overhang enhances the efficiency and specificity of cleavage, directing Dicer to cut ~22 nucleotides from the end of the substrate.

A number of previous studies have suggested that Dicer might function as an end-recognizing endonuclease, without positing a role for the 3' overhang. Processive Dicer cleavage was first implied by *in vitro* analysis of RISC cleavage

(*6*). In Drosophila embryo extracts programmed for RISC assembly using a long dsRNA, phased cleavage sites occurred at approximately 22 nucleotide intervals along an mRNA substrate. Similarly, analysis of *C. elegans* Dicer in whole cell extracts (*17*) or purified human Dicer *in vitro* (*35*) showed accumulation of discretely sized cleavage intermediates. Blocking of the ends of dsRNAs using either fold-back structures or chimeric RNA-DNA hybrids attenuated, but did not abolish, the ability of human Dicer to generate siRNAs (*35*). Finally, Lund and colleagues suggested that Dicer cleaved ~22 nt from the blunt end of an extended pre-miRNA, designed in part to mimic a pri-miRNA (see (*36*)).

Structural analysis of the PAZ domain suggested that it engages very short (~2-3 nucleotide) stretches of the 3' ends of single-stranded RNAs (*37-41*). This led Song and colleagues to propose a model in which the 3' overhang of pre-miRNAs served as an important recognition and specificity determinant for Dicer cleavage (*40*). Similarly, Dicer cleavage products would be preferential substrates for the enzyme as they also possess such overhangs. Our results are consistent with this model and suggest further that while the overhang is not obligate for Dicer processing of its substrates (see (*35*), and fig 1b), this structure does aid in determining the specificity of cleavage. Furthermore, time courses of processing of blunt and overhung 29nt shRNAs do show a more rapid processing of the overhung substrate if reactions are performed in the linear range for the enzyme (not shown).

To map more precisely the position of Dicer cleavage in the shRNA, we used primer extension analysis. The shRNAs described in Fig. 1A were reacted with recombinant human Dicer as shown in Figure 1B. Total RNA was recovered from the processing reactions and used in primer extension assays. Consistent with direct analysis of the RNA, shRNAs with 19 nt. stems failed to yield discrete extension products. The extension products that would be predicted from the unreacted substrate are not seen due to secondary structure of the uncleaved precursor (Fig 2A). Both of the 29 nt. shRNAs give rise to extension products with the overhung precursor giving a relatively discrete product of 20 nucleotides, as predicted for a cleavage precisely 22 nt. from the 3' end of the substrate (Fig 2b). The blunt-ended precursor gave a distribution of products, as was predicted from the analysis of uniformly and end-labeled RNAs.

In *Drosophila*, Dicer2 acts in a complex with a double-stranded RNA binding protein, R2D2 (*42*). Similarly, biochemical evidence from *C. elegans* suggests that its Dicer binds RDE-1, RDE-4 and DRH-1 (*43*). These results suggest that the human enzyme might also function as part of a larger complex, which could show altered cleavage specificities. Therefore, we also mapped the cleavage of our shRNAs *in vivo*. Precursors were transfected into cells, and the processed form of each was isolated by virtue of its co-immunoprecipitation with human Argonaute proteins, Ago1 and Ago2. Primer extension suggested identical cleavage specificities upon exposure of shRNAs to Dicer *in vitro* and in living cells (Fig 2c).

shRNAs are generally more effective than siRNAs

Since each shRNA gave rise to a single, predictable 22nt sequence in RISC, we could design an experiment to compare the efficacy of shRNAs and siRNAs. Toward this goal, we selected 43 sequences targeting a total of 6 genes (3-9 sequences per gene). For each sequence, we synthesized a 21 nt. siRNA (19 base stem) and 19 and 29 nt. shRNAs that were predicted to give Dicer products that were either identical to the siRNAs or that differed by the addition of one 3' nucleotide (Fig. 3a). Each RNA species was transfected into HeLa cells at a relatively high concentration (100 nM). The level of suppression was determined by semi-quantitative RT-PCR and the performance of each shRNA compared to the performance of the corresponding siRNA (Fig 3b). Comparison of 19 nt. shRNAs with siRNAs revealed that there was little difference in endpoint inhibition with these species (left panel). A comparison of siRNAs with 29 nt shRNAs gave a different result. Clustering of the comparison data points above the diagonal indicated consistently better endpoint inhibition with the 29 nt. shRNAs (right panel).

The generally better endpoint inhibition observed with 29 nt. shRNAs led us to investigate in more detail the performance of these silencing triggers as compared to siRNAs. Seventeen complete sets comprising an siRNA, a 19 nt. shRNA and a 29 nt. shRNA were examined for suppression in titration experiments. In all cases, the 19 nt. shRNAs performed as well as or worse than the corresponding siRNAs. In contrast, 29 nt. shRNAs exceeded the performance of siRNAs in the majority of cases. Four representative examples, targeting MAPK-14 are shown in Figure 3c. Several 29 nt. shRNAs (e.g., see MAPK14-1) showed both significantly greater endpoint inhibition and efficacy at lower concentrations than the corresponding siRNA. In other cases(e.g., see MAPK14-2 and MAPK-14-4), the maximal level of suppression for the 29 nt. shRNA was approximately two-fold greater than the maximal level of suppression for the corresponding siRNA. Finally, in a minority of cases, exemplified by MAPK14-3, the performance of the three types of silencing triggers was similar. Importantly, in only one case out of 17 did we note that the 29 nt. shRNA with a 2 nt. 3' overhang performed less effectively than the corresponding siRNA (data not shown).

siRNAs and shRNAs give similar profiles of off-target effects at saturation

Sequence specificity is a critical parameter in RNAi experiments. Microarray analysis has revealed down-regulation of many non-targeted transcripts following transfection of siRNAs into HeLa cells (*44*). Notably, these gene expression signatures differed between different siRNAs targeting the same gene. Many of the "off target" transcripts contained sites of partial identity to the individual siRNA, possibly explaining the source of the effects. To examine potential off-target effects of synthetic shRNAs, we compared shRNA signatures

with those of siRNAs derived from the same target sequence. Using microarray gene expression profiling, we obtained a genome-wide view of transcript suppression in response to siRNA and shRNA transfection. Figure 4(A and B) shows heat maps of signatures produced in HeLa cells 24 hours after transfection of 19 nt. and 29 nt. shRNAs compared with those generated by corresponding siRNAs. 19 nt. shRNAs produced signatures that resembled, but were not identical to, those of corresponding siRNAs. In contrast, the signatures of the 29 nt. shRNAs (Fig. 4B) were nearly identical to those of the siRNAs. These results indicate that off target effects may be inherent to the use of synthetic RNAs for eliciting RNAi and cannot be ameliorated by intracellular processing of an upstream precursor in the RNAi pathway. Furthermore, the agreement between the signatures of 29 nt. shRNAs and siRNAs is consistent with precise intracellular processing of the shRNA to generate a single siRNA rather than a random sampling of the hairpin stem by Dicer. The basis of the divergence between the signature of the 19 nt. shRNA and the corresponding siRNA is presently unclear.

Considered together, our results suggest that chemically synthesized, 29 nt. shRNAs are often substantially more effective triggers of RNAi than are siRNAs. A mechanistic explanation for this finding may lie in the fact that 29 nt. shRNAs are substrates for Dicer processing both *in vitro* and *in vivo*. We originally suggested that siRNAs might be passed from Dicer to RISC in a solid-state reaction on the basis of an interaction between Dicer and Argonaute2 in *Drosophila* S2 cell extracts (*9*). More recently, results from several laboratories have strongly suggested a model for assembly of the RNAi effector complex in which a multi-protein assembly containing Dicer and accessory proteins interacts with an Argonaute protein and actively loads one strand of the siRNA or miRNA into RISC (*32-34*). Such a model implies that Dicer substrates, derived from nuclear processing of pri-miRNAs or cytoplasmic delivery of pre-miRNA mimetics, might be loaded into RISC more effectively than siRNAs. Our data support such a prediction, since it is not the hairpin structure of the synthetic RNA that determines its increased efficacy but the fact that the shRNA is a Dicer substrate that correlates with enhanced potency. In *Drosophila*, Dicer is also required for siRNAs to enter RISC, and similar data has been obtained in mammalian cells (*32, 45*). Thus, it is possible that even siRNAs enter RISC via a Dicer-mediated assembly pathway and that our data simply reflect an increased affinity of Dicer for longer duplexes substrates. Alternatively, hairpin RNAs, such as miRNA precursors, might interact with specific cellular proteins that facilitate delivery of these substrates to Dicer, whereas siRNAs might not benefit from such chaperones. Overall, our results suggest an improved method for triggering RNAi in mammalian cells that uses higher potency RNAi triggers. Mapping the single 22nt sequence that appears in RISC from each of these shRNAs now permits the combination of this more effective triggering method with rules for effective siRNA design.

Methods

RNA sequence design

Each set of RNAs began with the choice of a single 19mer sequence. These 19mers were used directly to create siRNAs. To create shRNAs with 19mer stems, we appended a 4-base loop (either CCAA or UUGG) to the end of the 19mer sense strand target sequence followed by the 19mer complementary sequence and a UU overhang. To create 29mer stems, we increased the length of the 19mer target sequence by adding 1 base upstream and 9 bases downstream from the target region and used the same loop sequence and UU overhang. All synthetic RNA molecules used in this study were purchased from Dharmacon.

Dicer processing

RNA hairpins corresponding to luciferase were end-labeled with [γ-32P] ATP and T4 Polynucleotide kinase. 0.1 pmoles of RNA were then processed with 2 units of Dicer (Stratagene) at 37 °C for 2 hours. Reaction products were trizol extracted, isopropranol precipitated, run on an 18% polyacrylamide, 8M urea denaturing gel. For RNaseIII digestion, 0.1 pmoles were digested with 1 unit of E. coli RNase III (NEB) for 30 minutes at 37 °C and analyzed as described above. For primer extension analysis, hairpins were processed with Dicer at 37 °C for 2 hours, followed by heat inactivation of the enzyme. DNA primers were 5' labeled with PNK and annealed to 0.05 pmole of RNA as follows : 95 °C for one minute, 10 minutes at 50 °C and then 1 min on ice. Extensions were carried out at 42 °C for 1 hour using MoMLV reverse transcriptase. Products were analyzed by electrophoresis on a 8M Urea/20% polyacrylamide gel. For analysis of *in vivo* processing, LinxA cells were transfected in 10 cm plates using Mirus TKO (10 ug hairpin RNA) or Mirus LT4 reagent for DNA transfection (12 ug of tagged Ago1/Ago 2 DNA; J. Liu, unpublished). Cells were lysed and immunoprecipitated after 48 hours using with myc Antibody (9E14) Antibody. IPs were washed 3x in lysis buffer and treated with DNase for 15 minutes. Immunoprecipitates were then primer extended as described above.

siRNA and shRNA Transfections and mRNA Quantitation

HeLa cells were transfected in 96-well plates by use of Oligofectamine (Invitrogen) with the final nanomolar concentrations of each synthetic RNA indicated in the graphs. RNA quantitation was performed by Real-time PCR, using appropriate Applied Biosystems TaqMan™ primer probe sets. The primer probe set used for MAPK14 was Hs00176247_m1. RNA values were normalized to RNA for HGUS (probe 4310888E).

Microarray Gene Expression Profiling

HeLa cells were transfected in 6-well plates by use of Oligofectamine. RNA from transfected cells was hybridized competitively with RNA from mock-transfected cells (treated with transfection reagent in the absence of synthetic RNA). Total RNA was purified by Qiagen RNeasy kit, and processed as described previously (*46*) for hybridization to microarrays containing oligonucleotides corresponding to approximately 21,000 human genes. Ratio hybridizations were performed with fluorescent label reversal to eliminate dye bias. Microarrays were purchased from Agilent Technologies. Error models have been described previously (*46*). Data were analyzed using Rosetta Resolver™ software.

Acknowledgements

G.J.H. is supported by an Innovator Award from the U.S. Army Breast Cancer Research Program. This work was also supported by a grant from the NIH (GJH). We thank the Rosetta Gene Expression Laboratory for microarray RNA processing and hybridizations.

Figure Legends

Figure 1. *In vitro* processing of 29 nt. shRNAs by Dicer generates a single siRNA from the end of each short hairpin. **A.** The set of shRNAs containing 19 or 29 nt stems and either bearing or lacking a 2 nucleotide 3'overhang is depicted schematically. . For reference the 29 nt sequence from luciferase (top,blue) strand is given. The presumed cleavage sites are indicated in green and by the arrows. **B.** *In vitro* Dicer processing of shRNAs. Substrates as depicted in **A** were incubated either in the presence or absence of recombinant human Dicer (as indicated). Processing of a 500 bp. blunt-ended dsRNA is shown for comparison. Markers are end-labeled, single-stranded, synthetic RNA oligonucleotides. **C.** All shRNA substrates were incubated with bacterial RNase III to verify their double-stranded nature.

Figure 2. Primer extension analysis reveal a single siRNA generated from Dicer processing of shRNA both *in vitro* and *in vivo*. **A.** 19 nt. shRNAs, as indicated (see Fig 1A), were processed by Dicer *in vitro*. Reacted RNAs were extended with a specific primer that yields a 20 base product if cleavage occurs 22 bases from the 3' end of the overhung RNA (see Fig 1A). Lanes labeled siRNA are extensions of synthetic RNAs corresponding to predicted siRNAs that would be released by cleavage 21 or 22 nucleotides from the 3' end of the overhung precursor. Observation of extension products dependents entirely on the

inclusion of RT (indicated). Markers are phosphorylated, synthetic DNA oligonucleotides. **B.** Analysis as described in A for 29 nt. shRNAs. The * indicates the specific extension product from the overhung shRNA species. **C.** Primer extension were used to analyze products from processing of overhung 29 nt. shRNAs *in vivo*. For comparison, extensions of *in vitro* processed material are also shown. Again, the * indicates the specific extension product.

Figure 3
Gene suppression by shRNAs is comparable to or more effective than that achieved by siRNAs targeting the same sequences. **A.** Structures of synthetic RNAs used for these studies. **B.** mRNA suppression levels achieved by 43 siRNAs targeting 6 different genes compared with levels achieved by 19mer (left) or 29mer (right) shRNAs derived from the same target sequences. All RNAs were transfected at a final concentration of 100 nM. Values indicated on the *X* and *Y* axes reflect the percentage of mRNA remaining in HeLa cells 24 hours after RNA transfection compared with cells treated with transfection reagent alone. **C.** Titration analysis comparing efficacies of four siRNA/shRNA sets targeting MAPK14. Curves are graphed from data derived from transfections at 1.56, 6.25, 25, and 100 nM final concentrations of RNA. (Blue diamonds: 21mer siRNAs; pink squares: 19mer shRNAs; green triangles: 29mer shRNAs).

Figure 4. Microarray profiling reveals sequence-specific gene expression profiles and more similarity between 29mer shRNAs and cognate siRNAs than observed for 19mer shRNAs. Each row of the heat maps reports the gene expression signature resulting from transfection of an individual RNA. Data shown represent genes that display at least a 2-fold change in expression level (p value<0.01 and $\log_{10}$ intensity >1) relative to mock-transfected cells. Green indicates decreased expression relative to mock transfection whereas red indicates elevated expression. **A.** 19mer shRNAs and siRNAs designed for six different target sequences within the coding region of the MAPK14 gene were tested for gene silencing after 24 hours in HeLa cells. **B.** A similar experiment to that described in A but carried out with five 29mer shRNAs targeting MAPK14.

Supplementary Table 1. Sequences of the siRNAs used in this study

Gene	Accession number	Target sequence ID	Target sequence
IGF1R	NM_000875	IGF1R-1	GGAUGCACCAUCUUCAAGG
IGF1R	NM_000875	IGF1R-2	GACAAAAUCCCCAUCAGGA
IGF1R	NM_000875	IGF1R-3	ACCGCAAAGUCUUUGAGAA
IGF1R	NM_000875	IGF1R-4	GUCCUGACAUGCUGUUUGA
IGF1R	NM_000875	IGF1R-5	GACCACCAUCAACAAUGAG
IGF1R	NM_000875	IGF1R-6	CAAAUUAUGUGUUUCCGAA
IGF1R	NM_000875	IGF1R-7	CGCAUGUGCUGGCAGUAUA
IGF1R	NM_000875	IGF1R-8	CCGAAGAUUUCACAGUCAA
IGF1R	NM_000875	IGF1R-9	ACCAUUGAUUCUGUUACUU
KIF11	NM_004523	KIF11-1	CUGACAAGAGCUCAAGGAA
KIF11	NM_004523	KIF11-2	CGUUCUGGAGCUGUUGAUA
KIF11	NM_004523	KIF11-3	GAGCCCAGAUCAACCUUUA
KIF11	NM_004523	KIF11-4	GGCAUUAACACACUGGAGA

KIF11	NM_004523	KIF11-5	GAUGGCAGCUCAAAGCAAA
KIF11	NM_004523	KIF11-6	CAGCAGAAAUCUAAGGAUA
KIF14	NM_014875	KIF14-1	CAGGGAUGCUGUUUGGAUA
KIF14	NM_014875	KIF14-2	ACUGACAACAAAGUGCAGC
KIF14	NM_014875	KIF14-3	AAACUGGGAGGCUACUUAC
KIF14	NM_014875	KIF14-4	CACUGAAUGUGGGAGGUGA
KIF14	NM_014875	KIF14-5	GUCUGGGUGGAAAUUCAAA
KIF14	NM_014875	KIF14-6	CAUCUUUGCUGAAUCGAAA
KIF14	NM_014875	KIF14-7	GGGAUUGACGGCAGUAAGA
KIF14	NM_014875	KIF14-8	CAGGUAAAGUCAGAGACAU
KIF14	NM_014875	KIF14-9	CUCACAUUGUCCACCAGGA
KNSL1	NM_004523	KNSL1-1	GACCUGUGCCUUUUAGAGA
KNSL1	NM_004523	KNSL1-2	AAAGGACAACUGCAGCUAC
KNSL1	NM_004523	KNSL1-3	GACUUCAUUGACAGUGGCC
MAPK14	NM_139012	MAPK14-1	AAUAUCCUCAGGGGUGGAG
MAPK14	NM_139012	MAPK14-2	GUGCCUCUUGUUGCAGAGA
MAPK14	NM_139012	MAPK14-3	GAAGCUCUCCAGACCAUUU
MAPK14	NM_001315	MAPK14-4	CUCCUGAGAUCAUGCUGAA
MAPK14	NM_001315	MAPK14-5	GCUGUUGACUGGAAGAACA
MAPK14	NM_001315	MAPK14-6	GGAAUUCAAUGAUGUGUAU
MAPK14	NM_001315	MAPK14-7	CCAUUUCAGUCCAUCAUUC
PLK	NM_005030	PLK-1	CCCUGUGUGGGACUCCUAA
PLK	NM_005030	PLK-2	CCGAGUUAUUCAUCGAGAC
PLK	NM_005030	PLK-3	GUUCUUUACUUCUGGCUAU
PLK	NM_005030	PLK-4	CGCCUCAUCCUCUACAAUG
PLK	NM_005030	PLK-5	AAGAGACCUACCUCCGGAU
PLK	NM_005030	PLK-6	GGUGUUCGCGGGCAAGAUU
PLK	NM_005030	PLK-7	CUCCUUAAAUAUUUCCGCA
PLK	NM_005030	PLK-8	AAGAAGAACCAGUGGUUCG
PLK	NM_005030	PLK-9	CUGAGCCUGAGGCCCGAUA

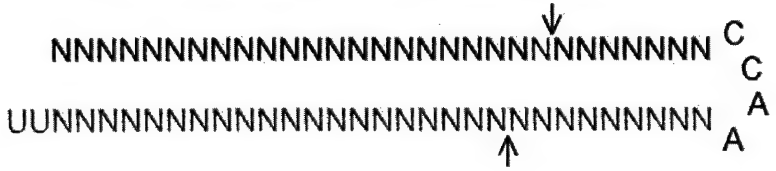
Literature Cited

1. A. Fire *et al.*, *Nature* **391**, 806-11. (Feb 19, 1998).
2. M. T. Ruiz, O. Voinnet, D. C. Baulcombe, *Plant Cell* **10**, 937-46. (Jun, 1998).
3. B. R. Williams, *Biochem Soc Trans* **25**, 509-13. (May, 1997).
4. G. J. Hannon, *Nature* **418**, 244-51. (Jul 11, 2002).
5. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950-2 (1999).
6. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25-33 (2000).
7. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **404**, 293-6 (2000).
8. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363-6. (Jan 18, 2001).
9. S. M. Hammond, S. Boettcher, A. A. Caudy, R. Kobayashi, G. J. Hannon, *Science* **293**, 1146-50. (Aug 10, 2001).
10. T. Tuschl, P. D. Zamore, R. Lehmann, D. P. Bartel, P. A. Sharp, *Genes Dev* **13**, 3191-7 (1999).
11. N. J. Caplen, S. Parrish, F. Imani, A. Fire, R. A. Morgan, *Proc Natl Acad Sci U S A* **98**, 9742-7. (Aug 14, 2001).
12. S. M. Elbashir *et al.*, *Nature* **411**, 494-8. (May 24, 2001).

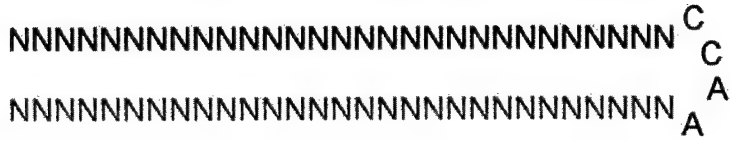
13. S. M. Elbashir, J. Martinez, A. Patkaniowska, W. Lendeckel, T. Tuschl, *Embo J* **20**, 6877-88. (Dec 3, 2001).
14. D. P. Bartel, *Cell* **116**, 281-97 (Jan 23, 2004).
15. Y. Lee *et al.*, *Nature* **425**, 415-9 (Sep 25, 2003).
16. G. Hutvagner *et al.*, *Science* **293**, 834-8. (Aug 3, 2001).
17. R. F. Ketting *et al.*, *Genes Dev* **15**, 2654-9. (Oct 15, 2001).
18. A. Grishok *et al.*, *Cell* **106**, 23-34. (Jul 13, 2001).
19. S. W. Knight, B. L. Bass, *Science* **293**, 2269-71. (Sep 21, 2001).
20. T. R. Brummelkamp, R. Bernards, R. Agami, *Science* **21**, 21 (2002).
21. P. J. Paddison, A. A. Caudy, E. Bernstein, G. J. Hannon, D. S. Conklin, *Genes Dev* **16**, 948-58. (Apr 15, 2002).
22. Y. Zeng, E. J. Wagner, B. R. Cullen, *Mol Cell* **9**, 1327-33. (Jun, 2002).
23. G. Sui *et al.*, *Proc Natl Acad Sci U S A* **99**, 5515-20. (Apr 16, 2002).
24. N. S. Lee *et al.*, *Nat Biotechnol* **20**, 500-5. (May, 2002).
25. C. P. Paul, P. D. Good, I. Winer, D. R. Engelke, *Nat Biotechnol* **20**, 505-8. (May, 2002).
26. R. C. Lee, V. Ambros, *Science* **294**, 862-4. (Oct 26, 2001).
27. N. C. Lau, L. P. Lim, E. G. Weinstein, D. P. Bartel, *Science* **294**, 858-62. (Oct 26, 2001).
28. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, *Science* **294**, 853-8. (Oct 26, 2001).
29. D. S. Schwarz *et al.*, *Cell* **115**, 199-208 (Oct 17, 2003).
30. J. M. Silva, R. Sachidanandam, G. J. Hannon, *Nat Genet* **35**, 303-5 (Dec, 2003).
31. A. Khvorova, A. Reynolds, S. D. Jayasena, *Cell* **115**, 209-16 (Oct 17, 2003).
32. Y. S. Lee *et al.*, *Cell* **117**, 69-81 (Apr 2, 2004).
33. J. W. Pham, J. L. Pellino, Y. S. Lee, R. W. Carthew, E. J. Sontheimer, *Cell* **117**, 83-94 (Apr 2, 2004).
34. Y. Tomari *et al.*, *Cell* **116**, 831-41 (Mar 19, 2004).
35. H. Zhang, F. A. Kolb, V. Brondani, E. Billy, W. Filipowicz, *Embo J* **21**, 5875-85. (Nov 1, 2002).
36. E. Lund, S. Guttinger, A. Calado, J. E. Dahlberg, U. Kutay, *Science* **303**, 95-8 (Jan 2, 2004).
37. J. B. Ma, K. Ye, D. J. Patel, *Nature* **429**, 318-22 (May 20, 2004).
38. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nat Struct Mol Biol* **11**, 576-7 (Jun, 2004).
39. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nature* **426**, 465-9 (Nov 27, 2003).
40. J. J. Song *et al.*, *Nat Struct Biol* **10**, 1026-32 (Dec, 2003).
41. K. S. Yan *et al.*, *Nature* **426**, 468-74 (Nov 27, 2003).
42. Q. Liu *et al.*, *Science* **301**, 1921-5 (Sep 26, 2003).
43. H. Tabara, E. Yigit, H. Siomi, C. C. Mello, *Cell* **109**, 861-71. (Jun 28, 2002).
44. A. L. Jackson *et al.*, *Nat Biotechnol* **21**, 635-7 (Jun, 2003).
45. N. Doi *et al.*, *Curr Biol* **13**, 41-6. (Jan 8, 2003).

46. T. R. Hughes *et al.*, *Nat Biotechnol* **19**, 342-7 (Apr, 2001).

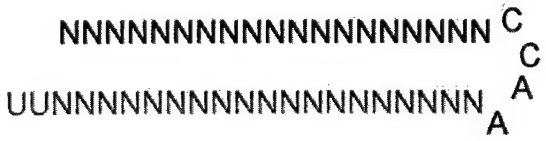
29 nt. shRNA with overhang



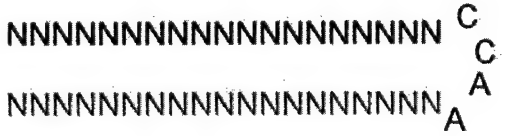
29mer shRNA no overhang



19 nt. shRNA with overhang



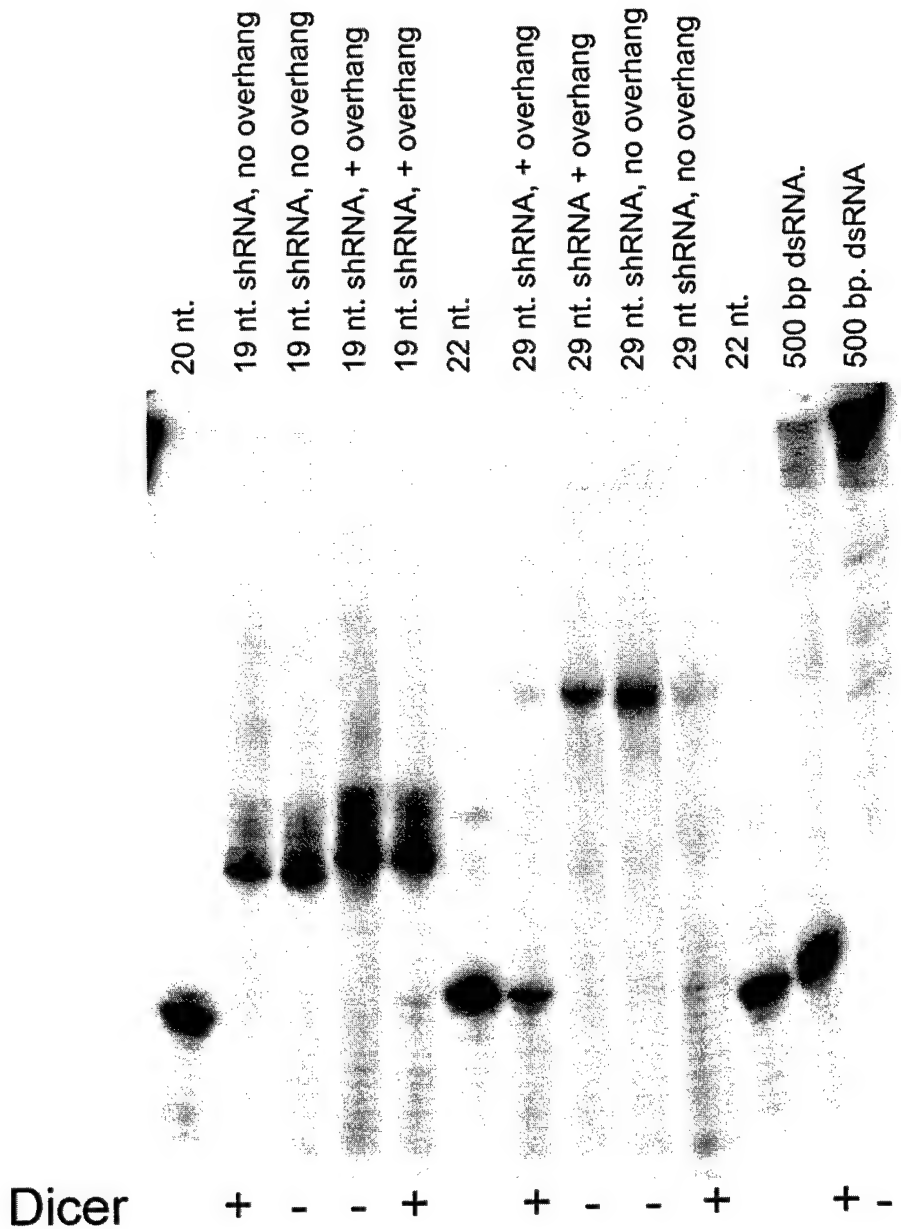
19mer shRNA



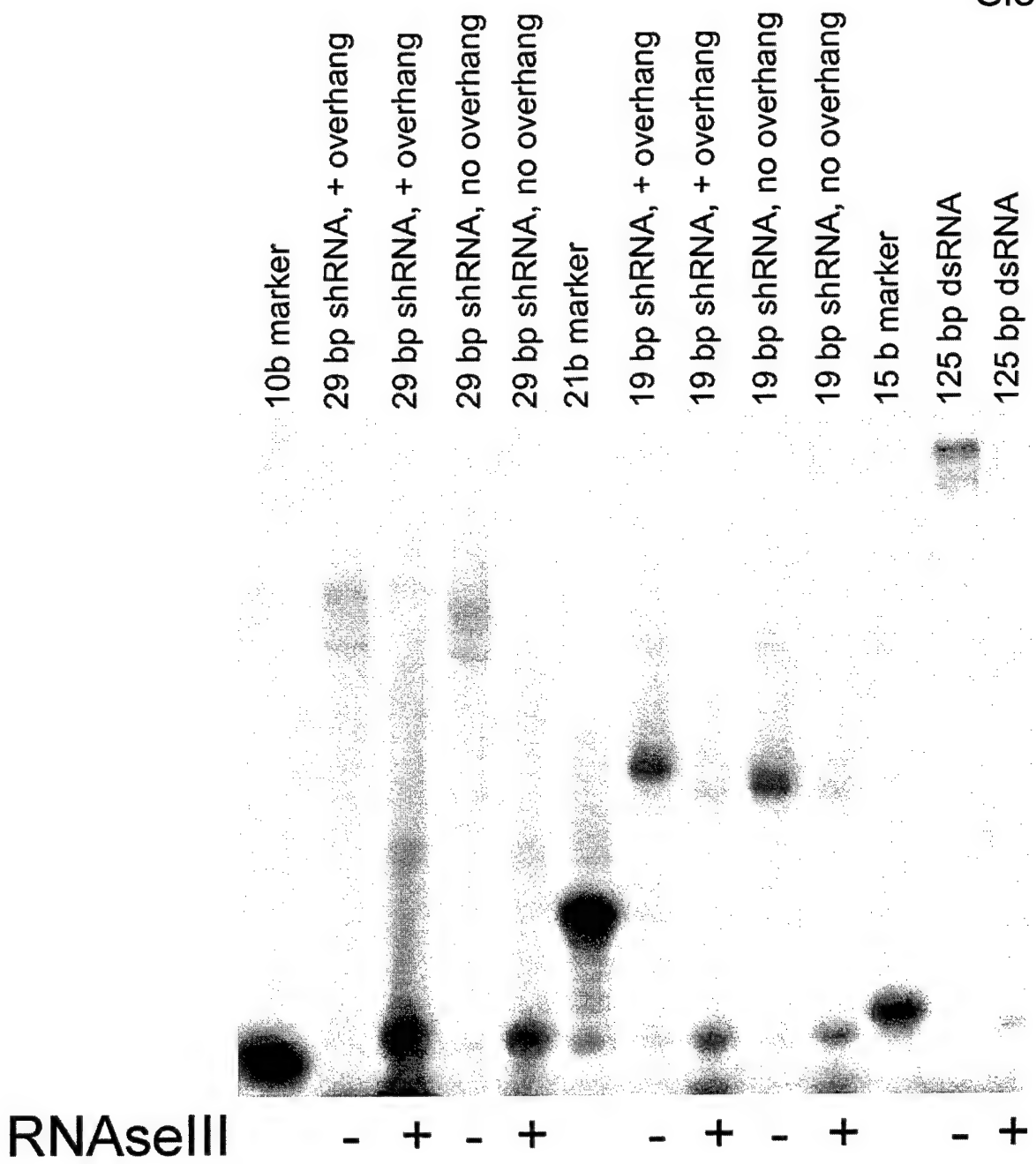
Luciferase 29mer

AGUUGCGCCCGCGAAUGAUAUUUAUAAUG

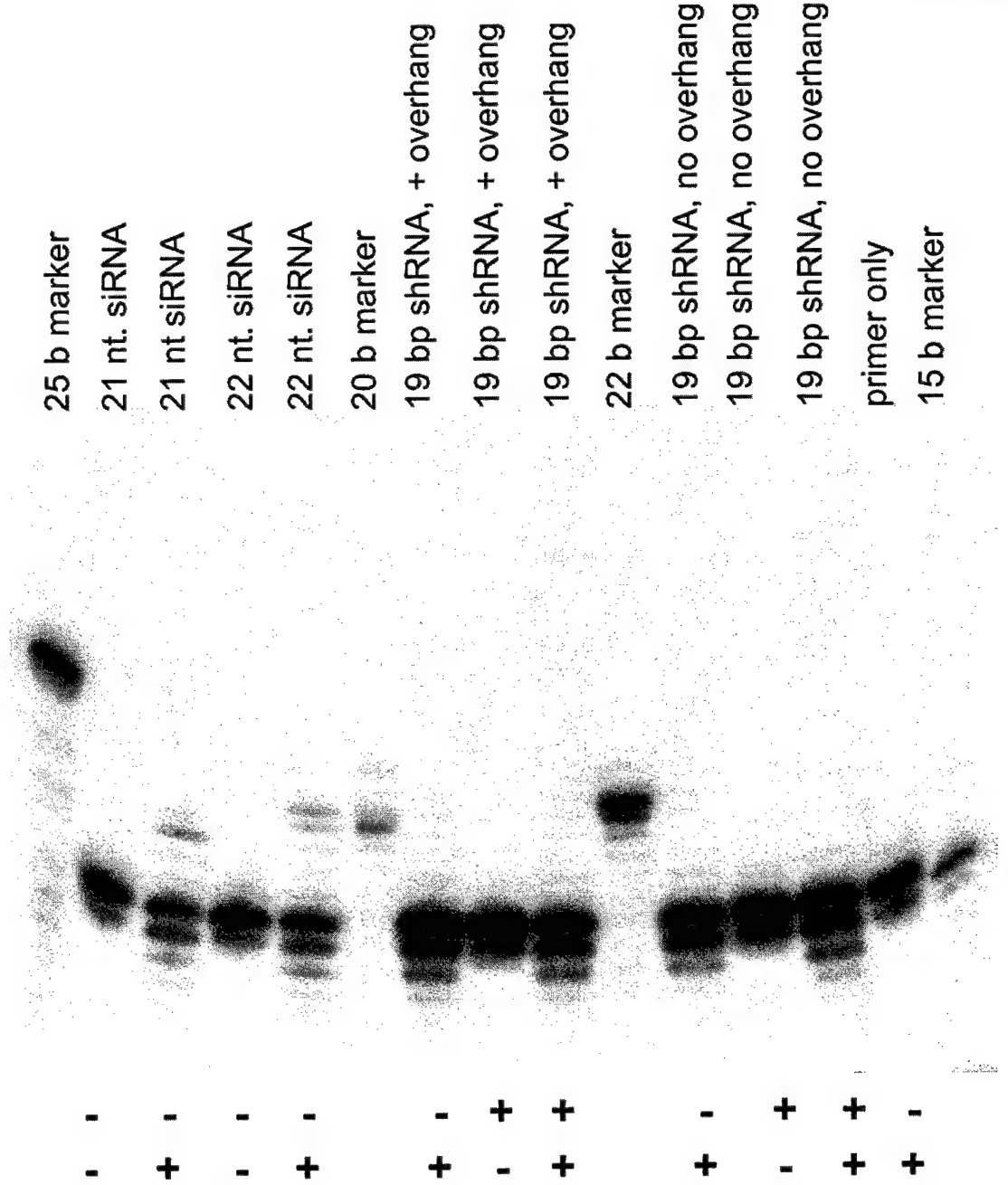
Siolas et al., Fig. 1b



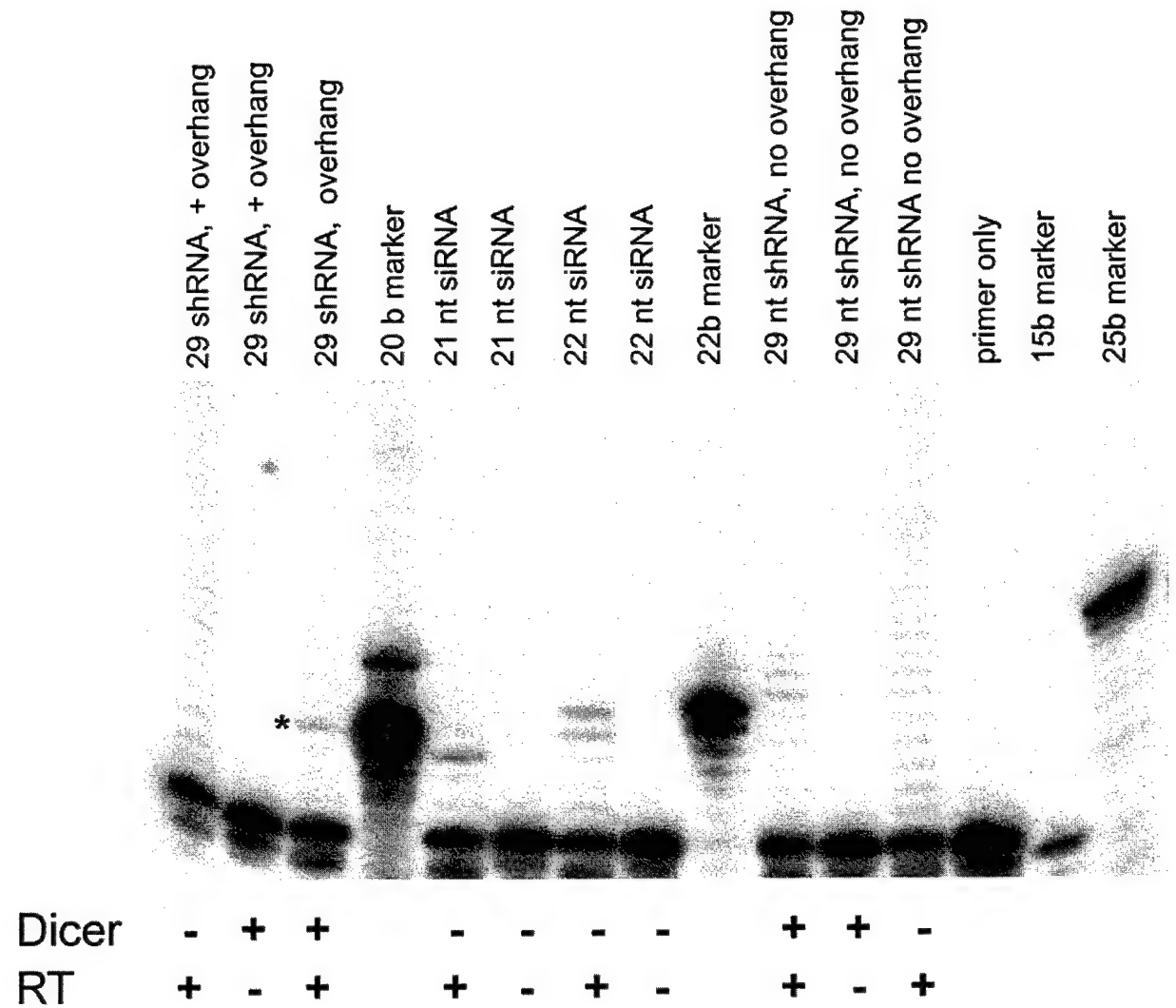
Siolas et al., Fig. 1c



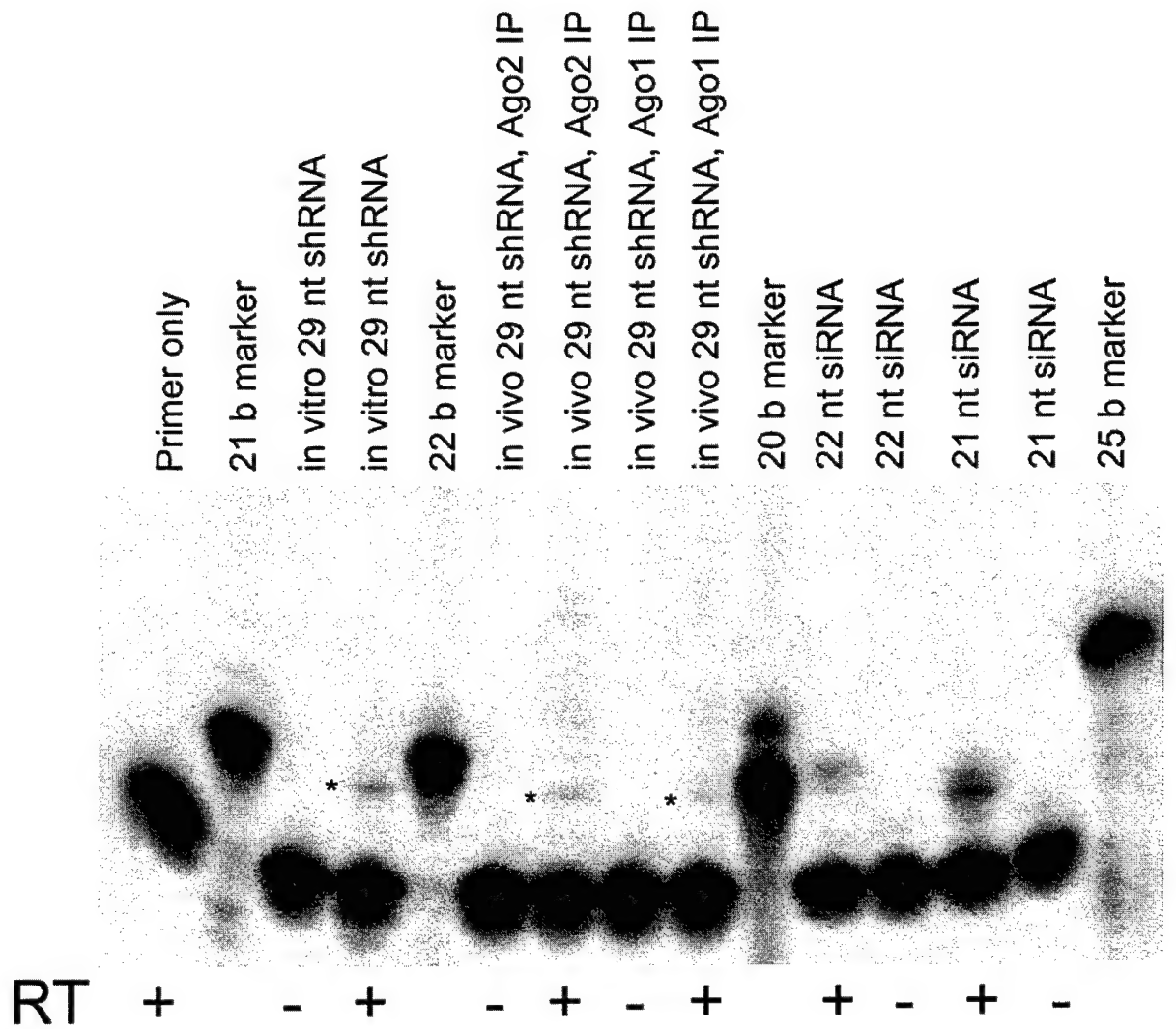
Siolas et al., Fig. 2a

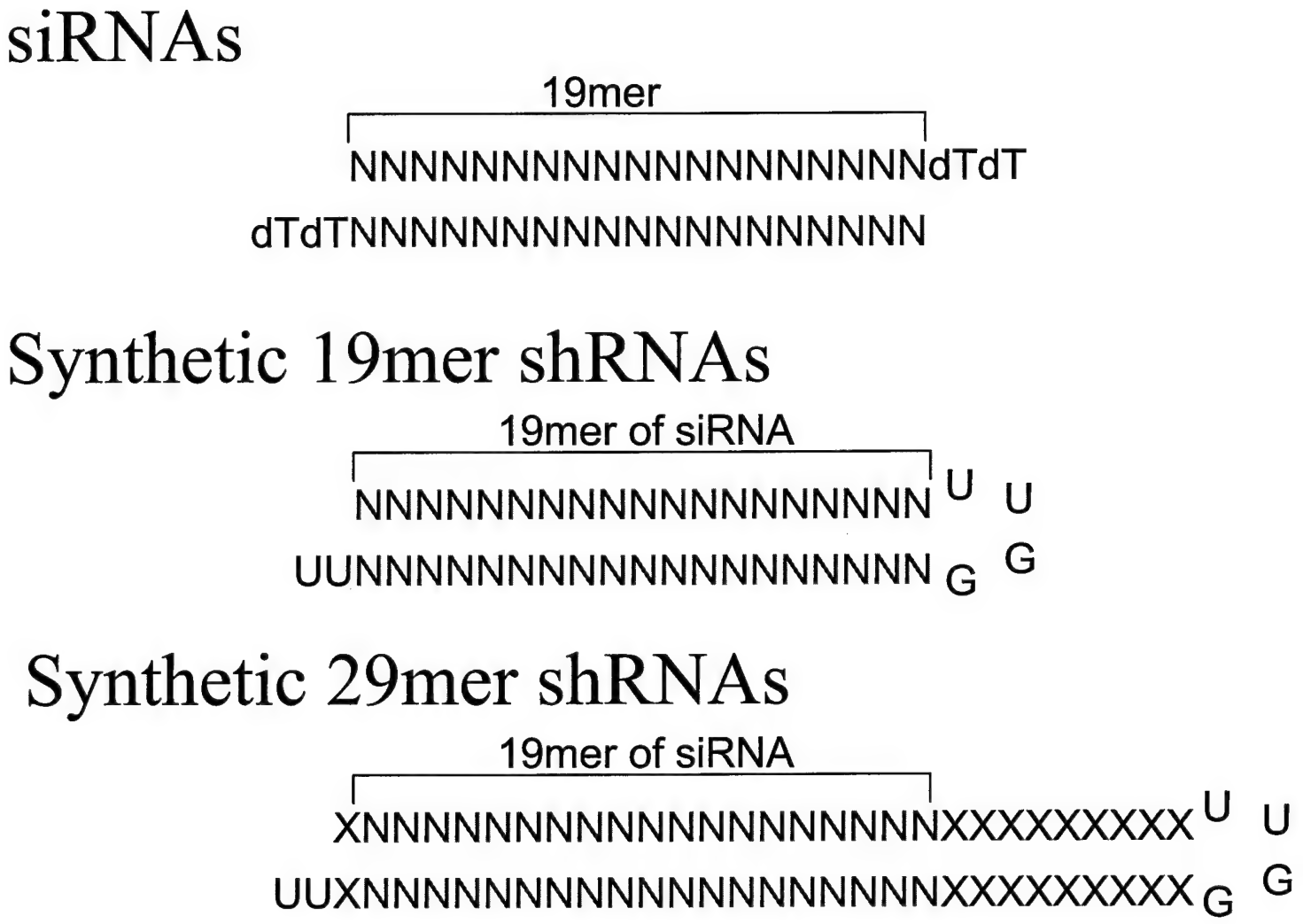


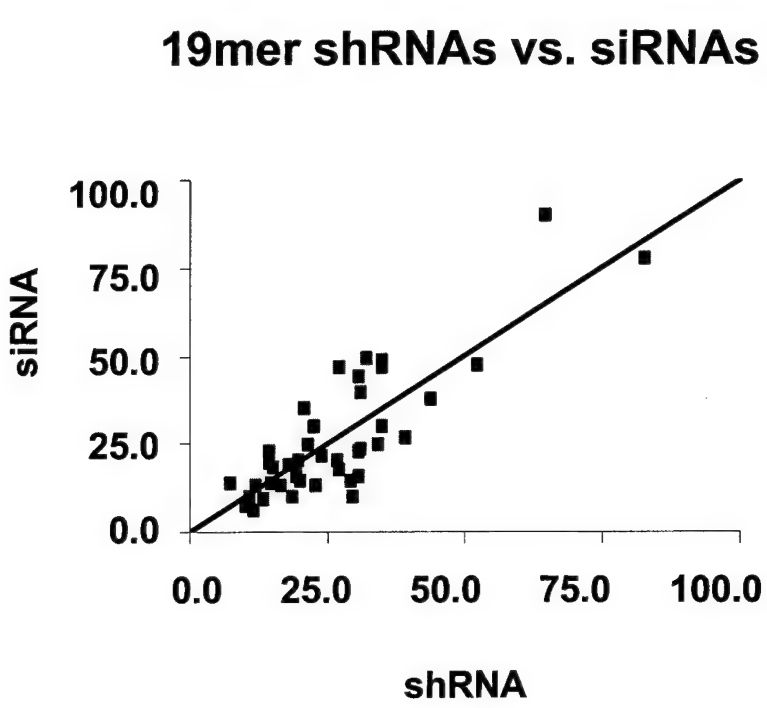
Siolas et al., Fig. 2b

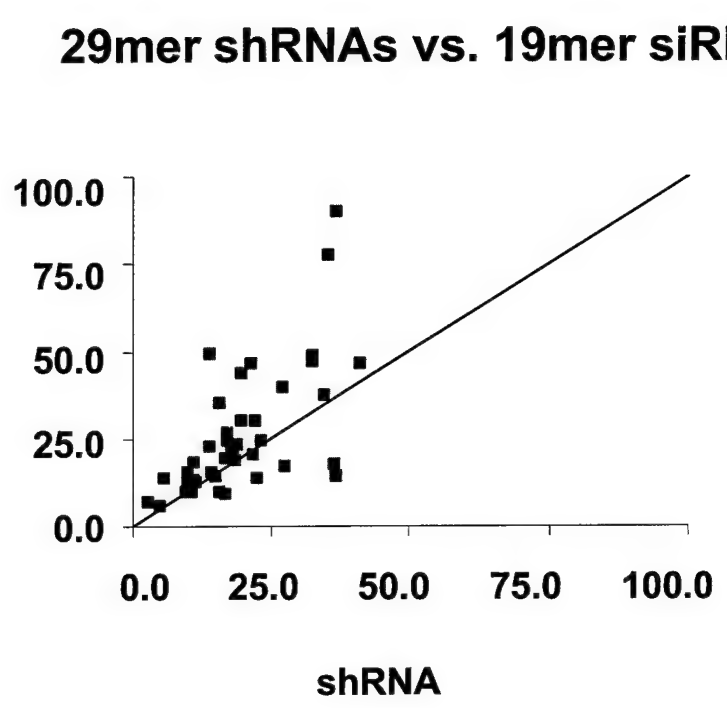


Siolas et al., Fig. 2c



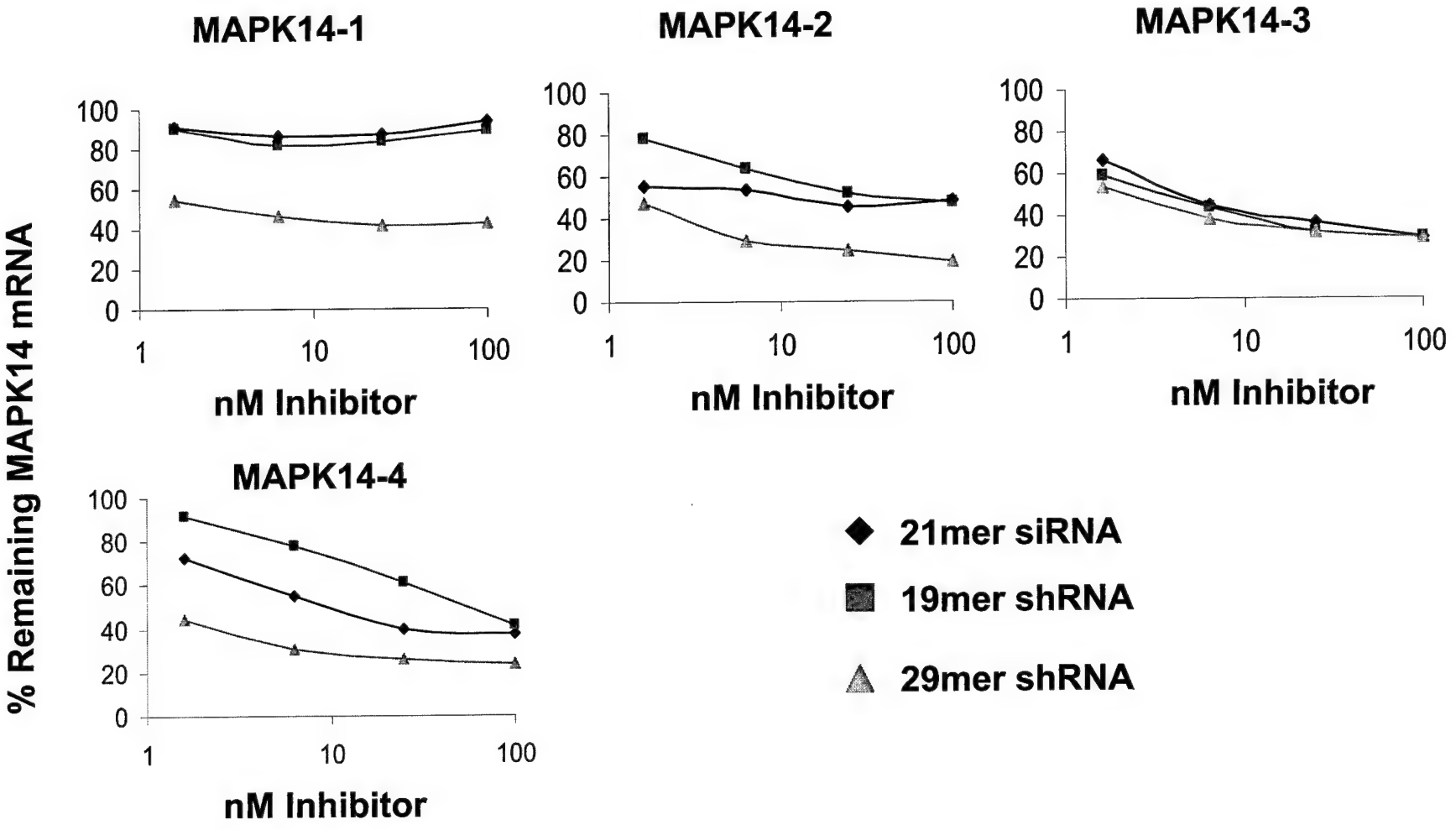

siRNAs
19mer
NNNNNNNNNNNNNNNNNNdTdT
dTdTNNNNNNNNNNNNNNNNNN
Synthetic 19mer shRNAs
19mer of siRNA
NNNNNNNNNNNNNNNNNNU U
UUNNNNNNNNNNNNNNNNNN G G
Synthetic 29mer shRNAs
19mer of siRNA
XNNNNNNNNNNNNNNNNNNXXXXXXXXXX U U
UUXNNNNNNNNNNNNNNNNNXXXXXXXXXX G G


19mer shRNAs vs. siRNAs
siRNA
100.0
75.0
50.0
25.0
0.0
0.0
25.0
50.0
75.0
100.0
shRNA

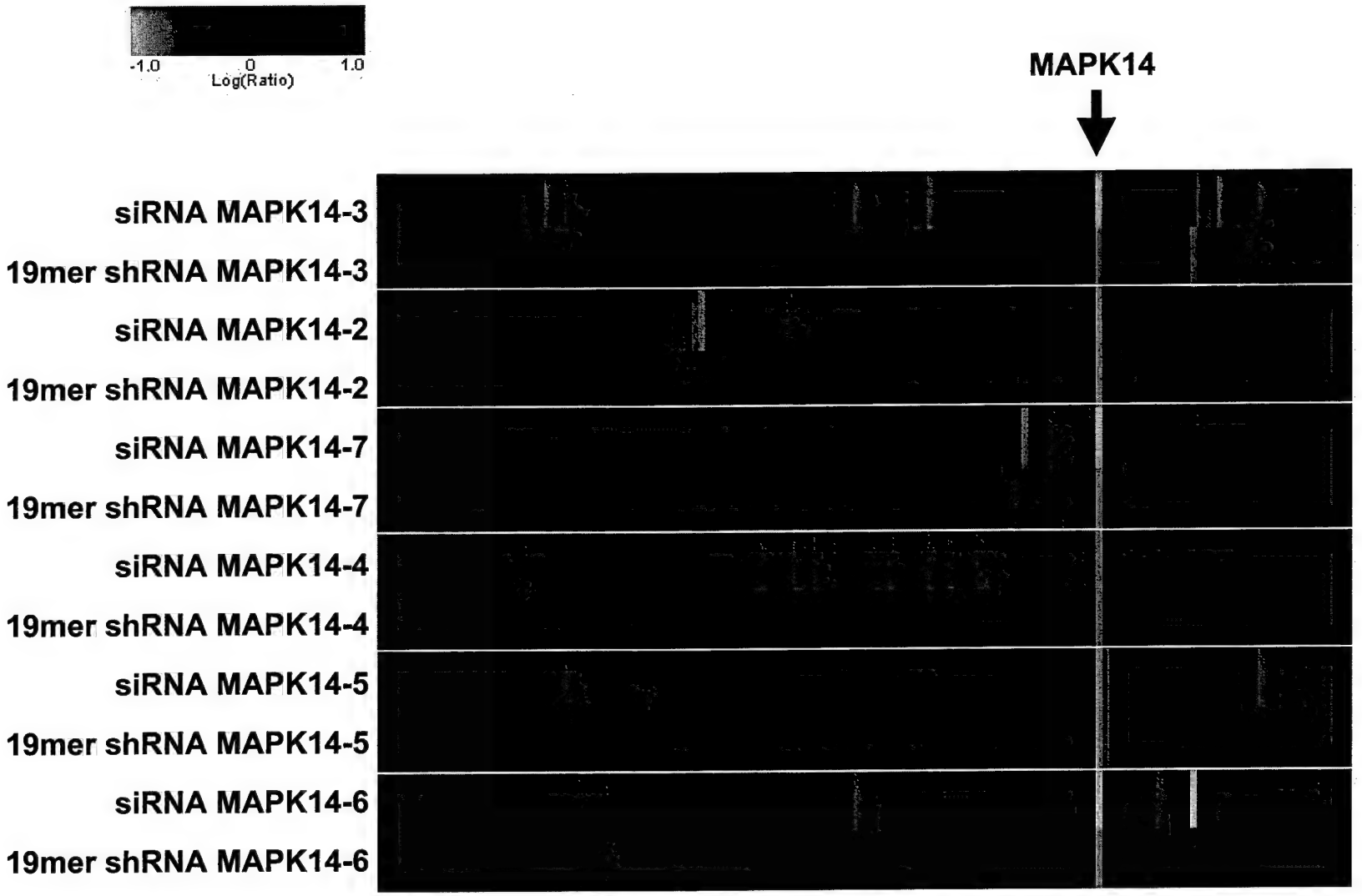

29mer shRNAs vs. 19mer siRNAs
100.0
75.0
50.0
25.0
0.0
0.0
25.0
50.0
75.0
100.0
shRNA

Siolas et al., Fig. 3c

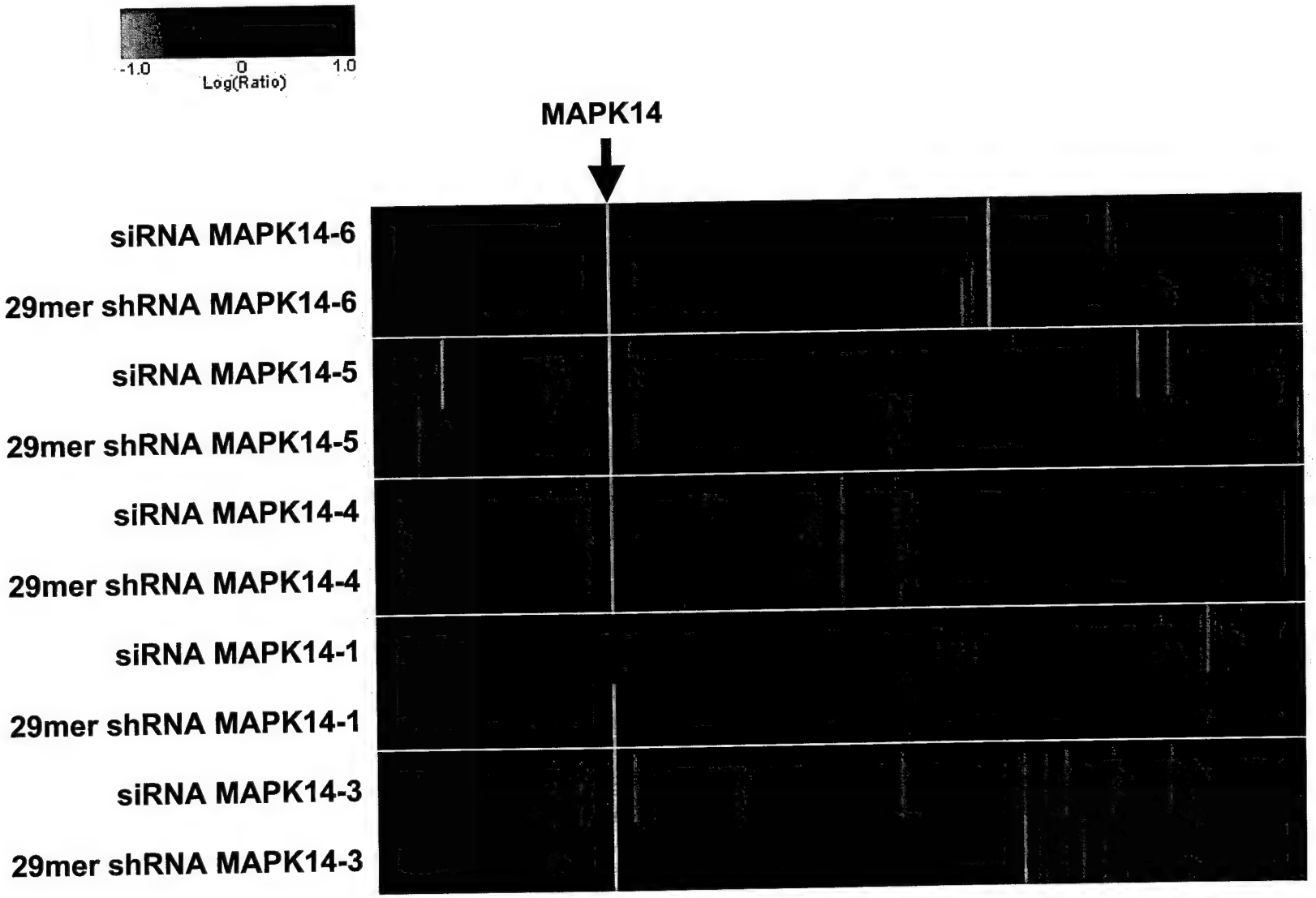
Figure 4C



Siolas et al., Fig. 4a



Siolas et al., Fig. 4b



A resource for large-scale RNA-interference-based screens in mammals

Patrick J. Paddison[1]*, Jose M. Silva[1]*, Douglas S. Conklin[1]*†, Mike Schlabach[2]†, Mamie Li[2], Shola Aruleba[1], Vivekanand Balija[1], Andy O'Shaughnessy[1], Lidia Gnoj[1], Kim Scobie[1], Kenneth Chang[1], Thomas Westbrook[2]†, Michele Cleary[3], Ravi Sachidanandam[1], W. Richard McCombie[1], Stephen J. Elledge[2]† & Gregory J. Hannon[1]

[1]*Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA*
[2]*Department of Biochemistry, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA*
[3]*Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA*

* These authors contributed equally to this work
† Present addresses: Department of Biomedical Sciences, Center for Functional Genomics, University at Albany, East Campus, B342A, One University Place, Rensselaer, New York 12144-2345, USA (D.S.C.); Department of Genetics, Harvard Partners Center for Genetics and Genomics, Harvard Medical School Room 158D, NRB, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA (M.S., T.W. and S.J.E.)

Gene silencing by RNA interference (RNAi) in mammalian cells using small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs) has become a valuable genetic tool[1–10]. Here, we report the construction and application of a shRNA expression library targeting 9,610 human and 5,563 mouse genes. This library is presently composed of about 28,000 sequence-verified shRNA expression cassettes contained within multi-functional vectors, which permit shRNA cassettes to be packaged in retroviruses, tracked in mixed cell populations by means of DNA 'bar codes', and shuttled to customized vectors by bacterial mating. In order to validate the library, we used a genetic screen designed to report defects in human proteasome function. Our results suggest that our large-scale RNAi library can be used in specific, genetic applications in mammals, and will become a valuable resource for gene analysis and discovery.

In invertebrates, RNAi has been harnessed as a powerful genetic tool to reduce endogenous gene expression through programming organisms or cells with homologous double-stranded (ds)RNA (reviewed in ref. 1). In *Caenorhabditis elegans*, this approach has been used in large-scale, genome-wide screens[2–5]. With the advent of RNAi in mammals and the refinement of techniques to trigger gene silencing, expression from any human or mouse transcript can, in principle, be inhibited using siRNAs or shRNAs.

To facilitate the use of RNAi as a genetic tool in mammals, we have constructed a large-scale library of RNAi-inducing shRNA expression vectors targeting human and mouse genes. We began by testing a number of variables that might affect shRNA performance. We previously showed that shRNAs containing 29 nucleotides of dsRNA and simple loop structures are effective silencing triggers when expressed from U6 small nuclear RNA promoters[6,7]. Other published accounts of shRNA-mediated inhibition include differences in promoter choice (reviewed in ref. 8), orientation of sense and antisense strands, length of RNA duplex (19–29 nucleotides), loop structure, and the addition of a 27-nucleotide U6 leader sequence[9]. After comparing each of these variables with a series of eight shRNAs targeting firefly luciferase or green fluorescent protein (GFP), we found that RNA polymerase III promoters were largely interchangeable (Supplementary Fig. 1a, b), that 29-nucleotide hairpins were more effective than shorter shRNAs (Supplementary Fig. 1a and data not shown), that changes in loop structure had minimal effects (data not shown), and that the addition of the U6 leader sequence had a positive effect (Supplementary Fig. 1c). The finalized hairpin design for this library is presented in Supplementary Fig. 2a. It contains a 27-nucleotide U6 leader sequence, followed by 29 base pairs (bp) of dsRNA and a 4-nucleotide loop. Recent studies have suggested that each shRNA is processed from its stem end by Dicer through a single cleavage event (D. Siolas and G.J.H., unpublished data). Thus, the combination of Drosha and Dicer processing is predicted to generate a precisely defined siRNA from each 29-nucleotide shRNA expression vector.

shRNAs were designed covering approximately 10,000 human and 5,000 mouse genes with between three and nine constructs each (Supplementary Fig. 2b). Only coding sequences were targeted, and each shRNA was chosen such that it contained >3 mismatches to any other gene. Where possible, shRNAs had sequence identity to the mouse orthologue of the targeted gene. In pilot experiments,

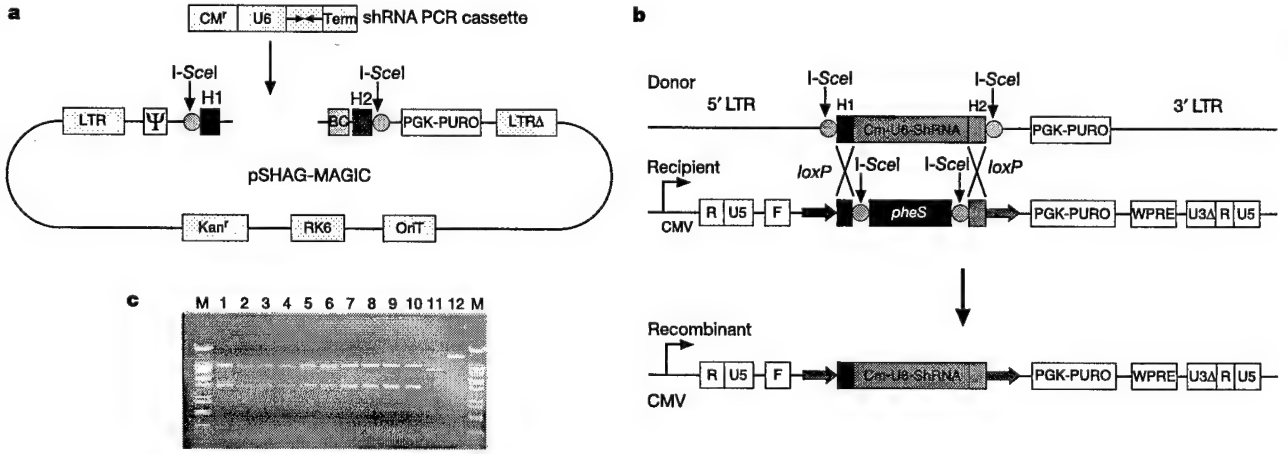


Figure 1 pSHAG-MAGIC shRNA cassette movement strategy. **a**, Map of the pSHAG-MAGIC vector. CMr, chloramphenicol-resistance gene; Kanr, kanamycin-resistance gene; LTR, long terminal repeat; OriT, origin of transfer. **b**, A diagrammatic representation of DNA exchanges occurring once the pSHAG-MAGIC donor vector has been transferred to cells containing a recipient vector by mating. In this case pLenti-LoxP (M.L. and S.J.E., unpublished data) is the recipient vector. WPRE, woodchuck hepatitis B post-transcriptional responsive element. **c**, Plasmids from ten independent colonies (post-mating) were digested with *NdeI* and the digestion products were separated on a 0.5% agarose gel (lanes 1–10). The parental plasmids, pSHAG-MAGIC (lane 11) and pLenti-LoxP (lane 12), each contain a single *NdeI* site, and on cleavage generate a 7.0- or 10.6-kilobase (kb) fragment, respectively. The pSHAG-MAGIC vector contains a unique *NdeI* site in the U6 promoter, which is transferred into the recipient vector. The correct mating product generates two fragments of 3.5 kb and 7.4 kb. M, marker.

between 25–75% of cloned shRNAs contained significant mutations, which arose during chemical synthesis. As a result we sequence-verified each shRNA in the library (see Supplementary Fig. 2b).

Our current sequence-verified library has a representation of 9,610 human and 5,563 mouse genes. This corresponds to 28,659 shRNAs targeting human genes and 9,119 shRNAs targeting mouse genes. Examples of library coverage for selected functional categories of human genes are presented in Supplementary Table 1. The most thoroughly represented functional groups in our current library are kinases and phosphatases, in which approximately 85% of all known kinases and phosphatases are targeted by three or more hairpins. Most other functional classes contain between 30–60% coverage with three or more hairpins, and >80% of genes in functional categories listed in Supplementary Table 1 are targeted by at least one sequence-verified hairpin.

The shRNA expression library has been constructed in a vector that contains a number of convenient design features (pSHAG-MAGIC, Fig. 1a). The vector is capable of producing self-inactivating murine-stem-cell-virus (MSCV) particles in commonly available retroviral packaging lines. We have also incorporated a new *in vivo* subcloning technology. This method is called 'Mating-Assisted Genetically Integrated Cloning' (MAGIC) (M.L. and S.J.E., manuscript in preparation). The MAGIC system consists of a donor vector (the library vector), in which the fragment of interest is flanked by two different 50-bp homology regions, H1 and H2, which in turn are flanked with linked I-*Sce*I sites. The donor vector also includes an F′ origin and a conditional origin of replication (RK6γ; Fig. 1a, b). The recipient vector, which also contains I-*Sce*I-linked H1 and H2 sites surrounding a negative selectable marker (*pheS*), resides in a bacterial strain that contains an inducible I-*Sce*I gene (Fig. 1b). After transfer of the donor vector into the recipient host by bacterial mating, I-*Sce*I cleaves both donor and recipient vectors, and these breaks are healed by homologous recombination via the H1 and H2 sequences. Selection against the unrecombined recipient containing *pheS* and I-*Sce*I sites and for the capture of the appropriate insert (chloramphenicol resistance) gives essentially 100% recovery of the desired plasmid. To test MAGIC transfer of shRNA clones, we developed a lentivirus recipient vector based on the FUW vector[10]. Using a test shRNA clone in the mating, we observed a mating efficiency of $>6 \times 10^6$ clones per ml of mixed bacteria. Restriction analysis showed that 10 out of 10 recombinants have the expected structure (Fig. 1c).

The shRNA library was designed to function for both genetic selections and screens. For selections conferring a growth advantage, the library can be used in a pooled fashion. Genetic screens (for example, for lethal events) can be carried out in a 96-well format; however, this method is labour intensive and time consuming. To facilitate such screens, we have adopted a DNA bar-coding strategy

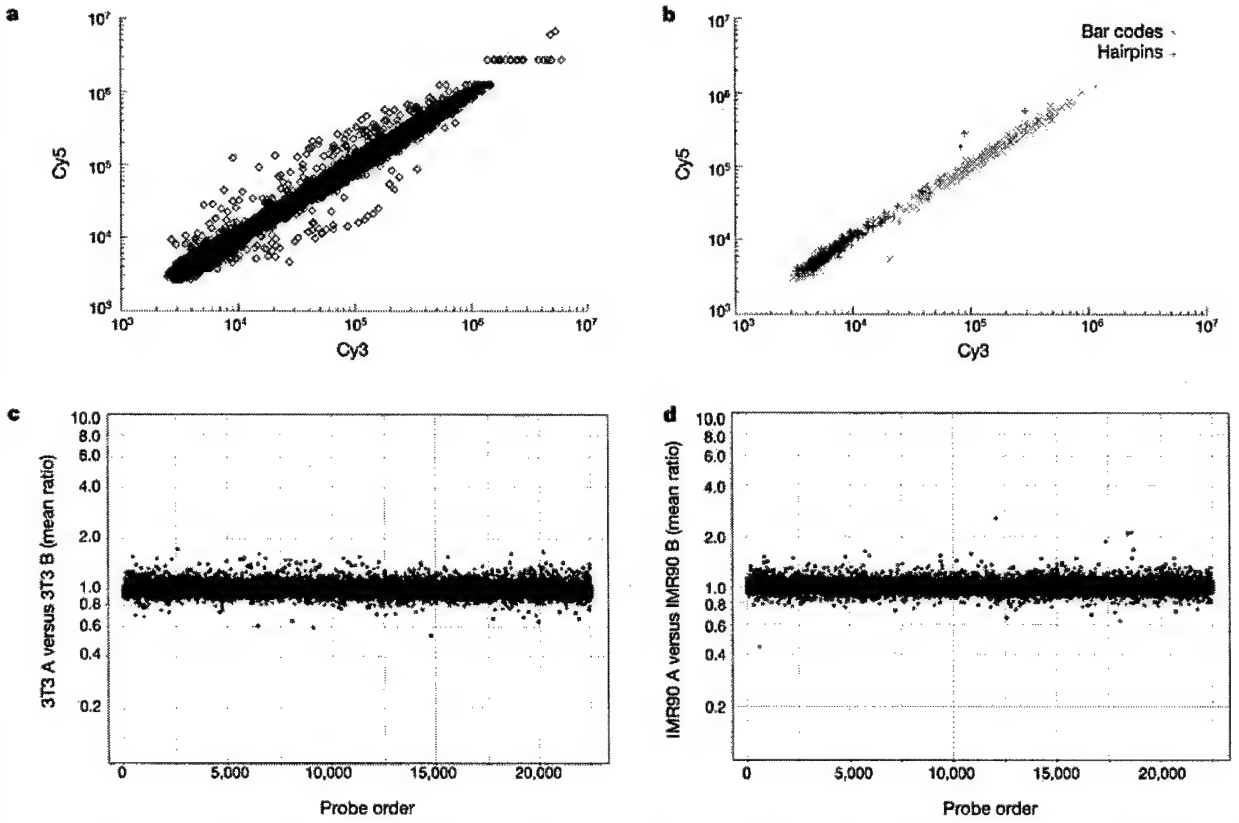


Figure 2 Microarray analysis of pSHAG-MAGIC library bar codes. **a**, Self–self hybridization of pSHAG-MAGIC library bar codes obtained from approximately 15,000 library plasmids prepared from *E. coli*. The DNA microarray (Agilent Technologies) is composed of 20,241 complementary 60-nuclotide oligonucleotides out of a total of 22,575 elements on the array, including controls; the remaining represent various positive and negative controls. Cy-labelled cRNA for library bar codes was generated and competitively hybridized as described in the Supplementary Methods. **b**, An analysis of a subset of 255 60-nucleotide bar codes from **a** versus 255 60-nucleotide shRNA probes. Each 60-nucleotide shRNA probe contains a direct repeat of the 29-nucleotide gene-targeting sequence with a 2-nucleotide spacer. **c**, Microarray analysis of pSHAG-MAGIC bar codes from transduced NIH3 T3 cells. Two pools of 2×10^7 cells (A and B) were infected separately and harvested 48 h after infection. Bar codes were processed and competitively hybridized as described in the Methods. The *x* axis is an index of all the elements (22,575), bar codes, hairpins and controls on the array. The *y* axis shows the mean measured ratios from two experiments (forward and reverse colour orientation) of 3T3 pool A hybridized against 3T3 pool B, plotted on a log scale with the numbers indicating fold change. Most probes, with the exception of a few, scatter around 1.0 (no change) with a range between 0.5 (twofold decrease) and 2.0 (twofold increase). **d**, Microarray analysis of pSHAG-MAGIC bar codes in transduced normal human diploid fibroblasts (IMR90) cells carried out as in **c**.

that has been used previously in *Saccharomyces cerevisiae* deletion collections, to follow individual mutants in complex populations via microarray analysis[11–13]. We have linked a unique 60-nuclotide DNA bar code to each shRNA vector to allow us to follow the fate of shRNAs in populations of virally transduced cells. To monitor relative frequencies of individual shRNA species, we designed microarrays (Agilent Technologies) containing 20,241 60-nucleotide DNAs complementary to each shRNA bar code.

Sequence analysis of library clones revealed that a fraction (about 4%) contained orphan shRNAs without bar codes. We therefore sought to determine whether we could use the shRNA sequence as a bar code. We chose 255 non-orphan shRNAs and deposited on arrays both associated bar codes and oligonucleotides designed have 60 hybridize to the shRNA itself. Such oligonucleotides were designed nucleotides in order to contain direct repeats of the 29-nucleotide antisense sequence. Cy-labelled antisense RNA was generated from about 15,000 library constructs such that it contained the reverse complement of the bar code and the entire shRNA. A linear log–log plot of a self hybridization revealed consistent cRNA production and hybridization from the plasmid library (Fig. 2a). However, hairpins gave substantially lower hybridization signals, essentially at background levels, relative to the corresponding bar codes (Fig. 2b), indicating that using hairpins as probes is not an optimal bar-coding strategy under these conditions.

We next asked whether bar codes could be used to report the representation of individual shRNAs in transduced mammalian cells. Normal human diploid fibroblasts (IMR90) or NIH3T3 cells

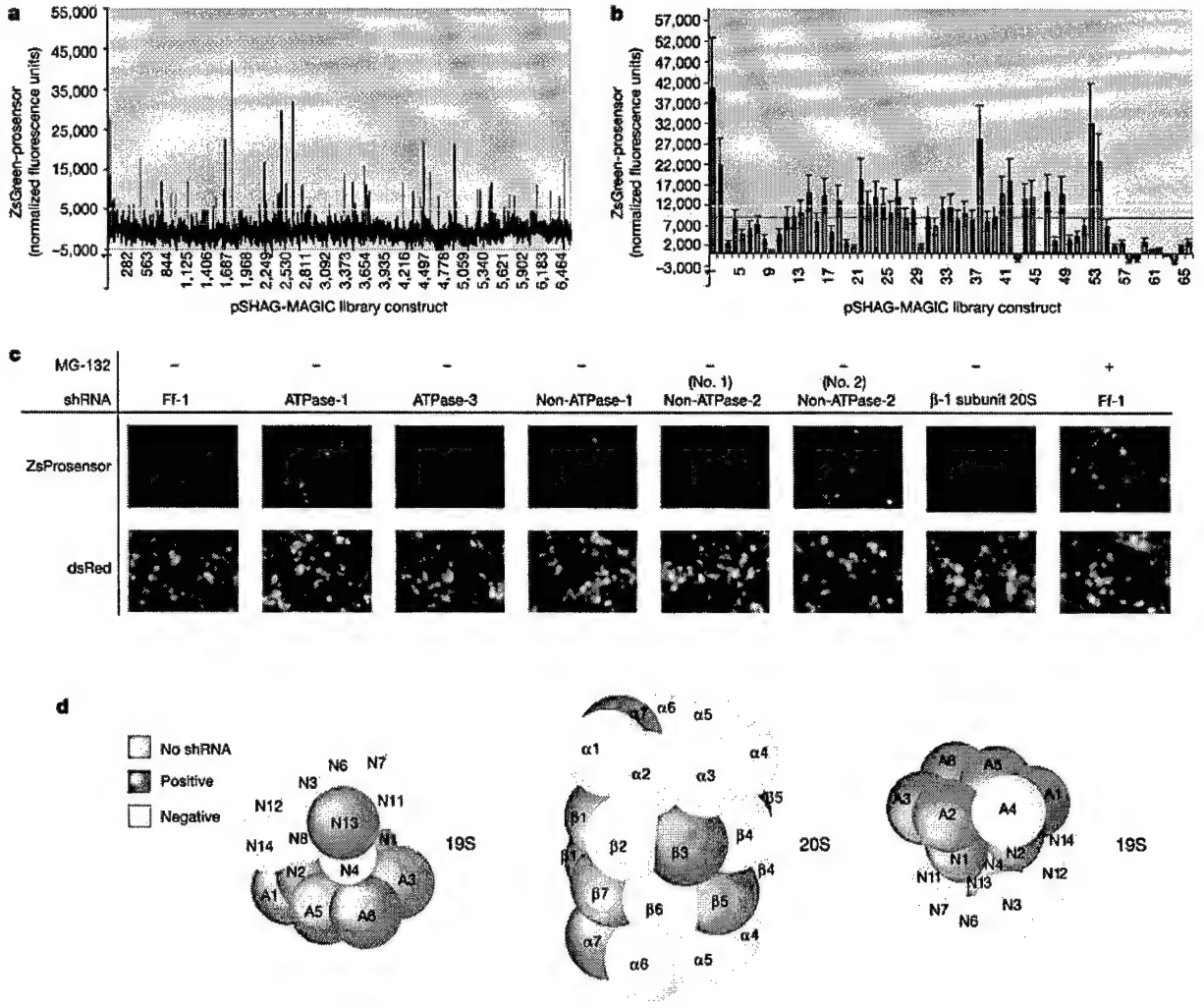


Figure 3 A reverse genetic screen for defects in human proteasome function. **a**, A graph of relative ZsGreen fluorescence for all pSHAG-MAGIC clones transfected. Red vertical bars indicate 22 positively scoring proteasome shRNAs corresponding to 15 known proteasome subunits, whereas black bars indicate library shRNAs. The yellow horizontal bar indicates a cut-off that was set based on control experiments. **b**, A replicate transfection experiment as in **a** using a set of 22 positively scoring proteasome hairpins and 33 that failed to score (blue bars), along with non-proteasome hairpin controls (green bars). These experiments were carried out in triplicate. All 22 of the proteasome hairpins with a positive score from the original screen were re-tested and also achieved a positive score. None of the 33 non-scoring proteasome hairpins from the original screen managed to match the score of the 22 positively scoring proteasome shRNAs from the first round (yellow line). However, in this experiment 36 of 55 scored well above (~2 s.d.) the mean background (red line). **c**, Fluorescence microscopy images showing representative results for individual proteasome hairpins as carried out in **a** and **b**. A shRNA targeting firefly luciferase (Ff-1) and small molecule proteasome inhibitor, MG-132 (Sigma), were used as negative and positive controls, respectively. **d**, A diagrammatic representation of the 26S proteasome colour-coded according to pSHAG-MAGIC library hits. Subunits coloured green had strong positive shRNA hits from the library in the primary screen. Grey subunits were not represented by any shRNAs in the approximately 7,000 tested. Subunits coloured blue were represented by at least 1 shRNA but did not score in the screen. For nomenclature see Supplementary Table 2.

were infected with retroviruses derived from the library population that was used for control hybridizations. To test for experimental variability, two populations of 2×10^7 cells (denoted A and B) were infected independently at a multiplicity of infection of approximately 1. Examination of colour-reversal experiments and plots comparing the relative intensity of the bar-code signals in these populations (Figs 2c, d) indicated that all steps of the procedure were highly reproducible, with the ratio of intensities varying by less than twofold in all but a few isolated cases for each cell line.

Finally, we sought to test the performance of the shRNA library in a biological context. We focused on an assay for which we could predict the recovery of a substantial number of shRNAs targeting genes in a known biological pathway. The 26S proteasome is the major non-lysosomal protease in eukaryotic cells. To search the library for shRNAs that compromise proteasome function, we used a reporter assay in which a fluorescent protein is coupled to a well-characterized degradation signal. The mouse ornithine decarboxylase (MODC) gene contains a PEST sequence that directs proteasomal degradation without the need for ubiquitination[14]. Roughly 7,000 shRNA expression plasmids, corresponding to approximately one-quarter of the complete library, were individually co-transfected with two expression constructs. The first encoded a *Zoanthus* green fluorescent (ZsGreen)–MODC degron fusion. In cases in which shRNAs compromised proteasome function, ZsGreen–MODC was expected to accumulate, giving a detectable signal. The second plasmid encoded *Discosoma* red fluorescent protein (DsRed). This permitted normalization of signals, which controlled for transfection efficiency.

An analysis of 6,712 shRNAs targeting 4,873 genes revealed approximately 100 RNAi constructs that increased the accumulation of ZsGreen–MODC (Fig. 3a). Twenty-two of these corresponded to 15 known proteasome subunits. As a secondary test, these 22 putative proteasome-positive shRNAs, and an additional 33 shRNAs that targeted proteasome subunits but which had not scored in the original screen, were selected from the population. These were assayed in replicate transfections for the ability to increase the stability of the unstable fluorescent protein. Again, the 22 shRNAs scored positively, whereas the 33 proteasome shRNAs not detected in the initial screen scored less well. However, an additional 14 shRNAs did score above background in the focused assay (Fig. 3b, c).

Notably, many of the positive shRNAs targeted 19S base subunits, including five out of five ATPases and the two largest non-ATPases (1 and 2) (Fig. 3c, d; see also Supplementary Table 2). Compared with the 19S base, targeting subunits of the 19S lid or the 20S core produced a lower hit rate. However, two out of three of the most important catalytic components, located in subunits β1 (peptidyl-glutamyl hydrolysing or caspase-like) and β5 (chymotrypsin-like)[15,16], achieved a positive score in our assay (Fig. 3). The other two β-subunits that we identified are involved in important *cis* contacts between neighbouring subunits (β3) and in *trans* contacts between the two β-rings (β7)[17,18]. In all cases tested, activation of the reporter was accompanied by a reduction in the expression of targeted proteasomal components (Fig. 4a and data not shown).

The c-Myc oncoprotein is a target for ubiquitin-mediated degradation by the proteasome[19]. As was observed for ZsGreen–MODC degron and bulk-ubiquitinated protein (not shown), c-Myc levels specifically increased upon transfection of cells with shRNAs that targeted proteasomal subunits (Fig. 4b).

Here, we report progress towards the construction of a genome-wide library of RNAi-inducing constructs for use in mammalian cells. At present the resource targets approximately 10,000 human and 5,000 murine genes, and it continues to expand. To examine the performance of the library, we have tested about one-quarter of its constituent clones (roughly 7,000 shRNA expression vectors) individually for the ability to inhibit the degradation of a direct proteasome target. Nearly 50% of the shRNAs in this collection that were expected to target proteosomal proteins were recovered as positives. The availability of this resource to the research community will open the door to the use of RNAi in mammalian systems as a large-scale tool for biological discovery. □

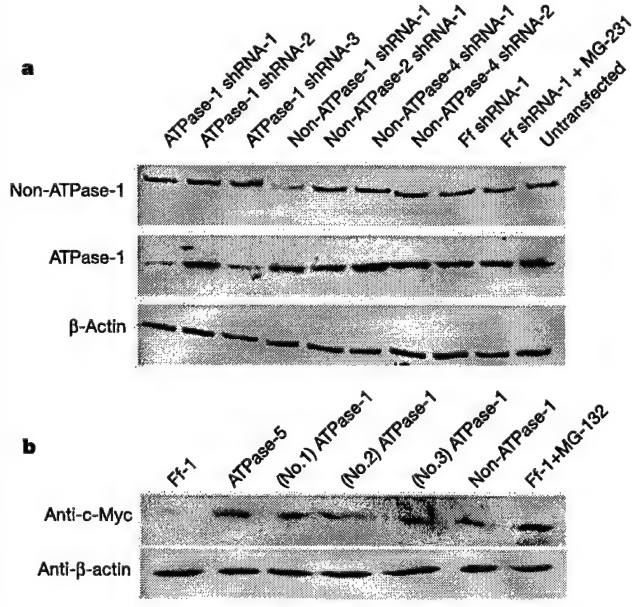


Figure 4 Further validation of selected pSHAG-MAGIC proteasome hairpins. **a**, Western blot showing specific suppression of the ATPase-1 and non-ATPase-1 of the 19S regulatory subunit in transiently transfected HEK293 cells. Cells were transfected with shRNAs as indicated. Knockdown of protein levels for shRNA-1 and -3 against ATPase-1 proteasome correlated with the severity of the relative scores in the pZsGreen–MODC accumulation assay. The lane labelled untransfected indicates the control lane where cells were not transfected. **b**, A western blot showing increased steady-state levels of endogenous c-Myc in HEK293 cells transiently transfected with library shRNAs as indicated. c-Myc is normally degraded by ubiquitin-mediated proteolysis in these cells[19].

Methods

A detailed protocol for the design and creation of shRNA polymerase chain reaction (PCR) cassettes is available at http://www.cshl.edu/public/SCIENCE/hannon.html. Oligonucleotides were purchased from Sigma-Genosys. The pSHAG-MAGIC shRNA library was cloned in the MAGIC-competent *Escherichia coli* strain BW23474F′DOT (M.L. and S.J.E., manuscript in preparation). The full sequence of the pSHAG-MAGIC vector and details of the mating protocol are available for download at the shRNA library website (http://www.cshl.edu/public/SCIENCE/hannon.html). Bar-code microarrays were synthesized by Agilent Technologies and were analysed as described in the Supplementary Methods.

Plasmids were prepared as described in the Supplementary Methods. Transfections were carried out in 96-well plates using plasmid mixtures described in Supplementary Methods. Fluorescence signals were read on a Victor2 plate reader. Signals in the green channel were normalized to transfection efficiency using customized scripts with fluorescence in the red channel serving as a normalization criterion. Cut-offs were assigned by using 16 independent control shRNA transfections to determine the range for a negative outcome.

Received 16 November 2003; accepted 26 January 2004; doi:10.1038/nature02370.

1. Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
2. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
3. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
4. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
5. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
6. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
7. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nature Genet.* **33**, 396–400 (2003).
8. Paddison, P. J. & Hannon, G. J. siRNAs and shRNAs: skeleton keys to the human genome. *Curr. Opin. Mol. Ther.* **5**, 217–224 (2003).

9. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nature Biotechnol.* **20,** 505–508 (2002).
10. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295,** 868–872 (2002).
11. Birrell, G. W., Giaever, G., Chu, A. M., Davis, R. W. & Brown, J. M. A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *Proc. Natl Acad. Sci. USA* **98,** 12608–12613 (2001).
12. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
13. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).
14. Ghoda, L., Sidney, D., Macrae, M. & Coffino, P. Structural elements of ornithine decarboxylase required for intracellular degradation and polyamine-dependent regulation. *Mol. Cell. Biol.* **12,** 2178–2185 (1992).
15. Chen, P. & Hochstrasser, M. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **86,** 961–972 (1996).
16. Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U. & Wolf, D. H. The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **272,** 25200–25209 (1997).
17. Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. & Huber, R. The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 295–317 (1999).
18. Coux, O. An interaction map of proteasome subunits. *Biochem. Soc. Trans.* **31,** 465–469 (2003).
19. Kim, S. Y., Herbst, A., Tworkowski, K. A., Salghetti, S. E. & Tansey, W. P. Skp2 regulates Myc protein stability and activity. *Mol. Cell* **11,** 1177–1188 (2003).

Supplementary Information accompanies the paper on **www.nature.com/nature.**

Acknowledgements We thank T. Moore and B. Simmons from Open Biosystems for their help in organizing and rearraying the library, and colleagues at CSHL and elsewhere (as indicated in Supplementary Table 1) as well as J. LaBaer and C. Perou for curating gene lists. G. Katari and J. Faith helped with bioinformatic analysis and shRNA choice, and members of the Lowe laboratory (CSHL) provided advice on vector optimization. This work was supported by an Innovator Award from the US Army Breast Cancer Research Program (G.J.H.), a contract from the National Cancer Institute (G.J.H.), grants from the NIH (G.J.H., W.R.M., S.J.E.) and the US Army Breast Cancer Research Program (G.J.H., D.S.C.), the Howard Hughes Medical Institute (S.J.E.), and by generous support from Oncogene Sciences and Merck. P.J.P. is an Arnold and Mabel Beckman Fellow of the Watson School of Biological Sciences and is supported by a predoctoral fellowship from the US Army Breast Cancer Research Program. J.M.S. is supported by a postdoctoral fellowship from the US Army Prostate Cancer Research Program. S.J.E. is an Investigator of the Howard Hughes Medical Institute. G.J.H. is a Rita Allen Foundation Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.org) or S.J.E. (selledge@genetics.med.harvard.edu).

A large-scale RNAi screen in human cells identifies new components of the p53 pathway

Katrien Berns[1]*, E. Marielle Hijmans[1]*, Jasper Mullenders[1], Thijn R. Brummelkamp[1], Arno Velds[1], Mike Heimerikx[1], Ron M. Kerkhoven[1], Mandy Madiredjo[1], Wouter Nijkamp[1], Britta Weigelt[2], Reuven Agami[3], Wei Ge[4], Guy Cavet[4], Peter S. Linsley[4], Roderick L. Beijersbergen[1] & René Bernards[1]

[1]*Division of Molecular Carcinogenesis and Center for Biomedical Genetics,* [2]*Division of Experimental Therapy, and* [3]*Division of Tumor Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*
[4]*Rosetta Inpharmatics, Inc., 12040 115th Avenue NE, Kirkland, Washington 98034, USA*

* These authors contributed equally to this work

RNA interference (RNAi) is a powerful new tool with which to perform loss-of-function genetic screens in lower organisms and can greatly facilitate the identification of components of cellular signalling pathways[1–3]. In mammalian cells, such screens have been hampered by a lack of suitable tools that can be used on a large scale. We and others have recently developed expression vectors to direct the synthesis of short hairpin RNAs (shRNAs) that act as short interfering RNA (siRNA)-like molecules to stably suppress gene expression[4,5]. Here we report the construction of a set of retroviral vectors encoding 23,742 distinct shRNAs, which target 7,914 different human genes for suppression. We use this RNAi library in human cells to identify one known and five new modulators of *p53*-dependent proliferation arrest. Suppression of these genes confers resistance to both *p53*-dependent and $p19^{ARF}$-dependent proliferation arrest, and abolishes a DNA-damage-induced G1 cell-cycle arrest. Furthermore, we describe siRNA bar-code screens to rapidly identify individual siRNA vectors associated with a specific phenotype. These new tools will greatly facilitate large-scale loss-of-function genetic screens in mammalian cells.

RNAi is a defence mechanism triggered by double-stranded (ds)RNAs to protect cells from parasitic nucleic acids. The dsRNAs are processed into siRNAs, which target homologous RNAs for destruction[6]. In mammalian cells, an RNAi response can be triggered by 21-base-pair siRNAs, which can cause strong, but transient, inhibition of gene expression[7]. By contrast, vector-expressed shRNAs can suppress gene expression over prolonged periods[4,5]. In the current study, we create a large set of vectors and use them to search for components of the p53 tumour-suppressor pathway. This pathway is crucial for genome integrity as it transmits both anti-proliferative and pro-apoptotic signals in response to a variety of stress signals[8].

To construct a human RNAi library (the 'NKi library'), we selected 7,914 human genes for shRNA-mediated reduction in expression, known as knockdown. This collection of genes includes components of major cellular pathways, including the cell cycle, transcription regulation, stress signalling, signal transduction and important biological processes such as biosynthesis, proteolysis and metabolism. In addition, genes implicated in cancer and other diseases are included in the library (see Supplementary Table 1). To increase the likelihood of obtaining a significant inhibition of gene expression, we constructed three different shRNA vectors against each gene (23,742 vectors in total; Fig. 1a and Supplementary Fig. 1). The oligonucleotides specifying the shRNAs were annealed and cloned in a high-throughput fashion into pRetroSuper (pRS), a retroviral vector that contains the shRNA expression cassette[9]. Using a pool of three knockdown vectors against a single gene, we obtain on average 70% inhibition of expression for approximately 70% of the genes in the library (see ref. 10; data not shown). The vector-based shRNA library can be used for functional genetic screens in both short-term and long-term assays using DNA transfection or retroviral transduction.

To validate the RNAi library, we developed a cell system to screen for bypass of p53-dependent proliferation arrest. We generated primary human BJ fibroblasts, which ectopically express the murine ecotropic receptor, the telomerase catalytic subunit (TERT) and a temperature-sensitive allele of SV40 large T antigen (tsLT), yielding BJ-TERT-tsLT cells. As can be seen in Fig. 1b, these cells proliferate when grown at 32 °C, the temperature at which tsLT binds and inactivates both retinoblastoma protein (pRB) and p53, but enter into a synchronous proliferation arrest after a shift to 39 °C, at which tsLT is inactive. To determine whether this proliferation arrest is p53 dependent, we infected BJ-TERT-tsLT cells with pRS–p53 (which targets *p53* for suppression) at 32 °C, and shifted them to 39 °C after two days. Figure 1b shows that knockdown of *p53* allowed temperature-shift-induced proliferation arrest to be bypassed. Knockdown of the RB pathway component $p16^{INK4A}$ alone did not allow growth arrest to be bypassed, but simultaneous suppression of both $p16^{INK4A}$ and *p53* yielded a further increase in escape from growth arrest compared with knockdown of *p53* alone (Fig. 1b). Thus, the conditional proliferation arrest in BJ fibroblasts depends primarily on *p53*.

We isolated polyclonal plasmid DNA from each of the 83 pools of

APPENDIX 3

Production of complex libraries of defined nucleic acid sequences using highly parallel *in situ* oligonucleotide synthesis

Michele Cleary[1*], Kristopher Kilian[1], Yanqun Wang[1], Jeff Bradshaw[1], Guy Cavet[1], Wei Ge[1], Amit Kulkarni[1], Patrick J. Paddison[2], Kenneth Chang[2], Nihar Sheth[2], Eric Leproust[3], Ernest M. Coffey[1], Julja Burchard[1], Peter Linsley[1] and Gregory J. Hannon[2*]

[1]Rosetta Inpharmatics LLC,
A Wholly Owned Subsidiary of Merck & Co., Inc.
401 Terry Ave North
Seattle, Washington 98109, USA

[2]Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, New York 11724, USA

[3]Agilent Technologies
3500 Deer Creek Road
Palo Alto, California 94304

*To whom correspondence should be addressed
Michele_Cleary@merck.com
hannon@cshl.edu

The ability to generate complex libraries of defined nucleic acid sequences can greatly aid in the functional analysis of genomes and in the study of nucleic acid interaction sites for a variety of proteins. Previously, such studies depended on either synthetic oligonucleotides or cellular nucleic acids as a starting material. Each of these methods has particular disadvantages. Therefore, we developed a rapid and cost-effective technique for construction of small fragment DNA libraries of defined sequences. This approach uses ink-jet DNA synthesis for generation of complex oligonucleotide populations *in situ* on microarray slides. These populations can be recovered and either used directly or immortalized by cloning. From a single microarray, a library containing thousands of unique sequences can be generated. We have validated this approach for the production of plasmids encoding short hairpin RNAs (shRNAs) targeting numerous human and mouse genes. This approach results in high fidelity clone retrieval with a uniform representation of the intended library sequences.

Introduction

Nucleic acid libraries provide some of the most versatile tools for functional analysis of genomes or of individual proteins or complexes. For example, cDNA expression libraries have been used to decipher biological pathways through the analysis of genetic interactions. By use of such tools gain-of-function (ectopic expression) experiments have revealed regulators of the p53 pathway [1,2] and apoptosis[3-5] and have permitted the cloning of cellular receptors for secreted signaling molecules [6-8] and viruses [9,10]. Additionally, antisense RNA expression libraries have been used to decipher elements of tumor suppressor function [11], invasion and metastasis [12] through loss-of-function genetics. In each of these approaches, libraries suffer from failing to cover all expressed sequences in the genome, largely due to tissue-specific mRNA expression and variations in mRNA abundance limiting the complexity and uniform representation of the cDNA source material.

More recently, the ability to harness loss-of-function genetics in many organisms has improved greatly due to the application of technologies based on RNA interference [13]. Indeed for several systems, including *Drosophila*, *C. elegans* and mammals, progress toward the construction of genome-wide libraries of RNAi-inducing DNA constructs has been substantial [14-23]. In the case of mammalian systems, two approaches have been taken. Several groups relied on endogenous nucleic acids, converted into cDNAs, as the source material for the construction of short hairpin libraries [20-23]. Our group and that of Bernards used conventional oligonucleotide synthesis to construct libraries of defined shRNA sequences [18,19]. The latter approach has several advantages. First is the generation of a pool of sequences with a uniform representation. Additionally, sequences within the defined library can be chosen to maximize

efficiency on the basis of rules for choosing effective RNAi triggers, while simultaneously minimizing off-target effects that are caused by cross-hybridization of shRNAs to multiple mRNAs [24-26]. Each sequence can be fully verified by sequence analysis, which also allows the establishment of linkage between a given sequence and a molecular tag or barcode [18,19]. Of course, the disadvantage of this approach is the high cost of producing large numbers of DNA oligonucleotides by conventional synthesis. This ultimately limits both the coverage of the resulting libraries and the range of model organisms for which they can be generated.

To address the cost inherent in the use of conventional methods for the generation of complex libraries of defined nucleic acids, we developed an approach that relies on the use of printed microarrays as a source material for complex oligonucleotide populations [27-30] (see Fig. 1). The oligonucleotides are designed with common sequences at their 5' and 3' ends that serve as PCR primer binding sites. For vector library construction, the cleaved material can be amplified and the products cloned into a plasmid of choice. Alternatively, the harvested oligonucleotides or the pool of PCR products could themselves constitute a library. We have validated this approach for the construction of libraries of shRNA expression constructs covering all of the predicted genes in the mouse and human genomes.

Results

Ink-jet technology has been optimized for hybridization microarrays using oligonucleotides of 60 nucleotides or less on slides that contain ~25,000 individual spots [27-30]. However, no definitive tests have allowed predictions of either the fidelity of synthesis or the overall amount of intact oligonucleotide product that is generated via this method. To test the feasibility of using microarrays to produce nucleic acid libraries, we first designed and printed arrays containing 110 unique 59-nt DNA sequences, each containing identical PCR primer binding sites. To increase the amount of available PCR template, each oligonucleotide was synthesized redundantly in ~220 different locations. Oligonucleotide populations were recovered from the microarray surface using either of two approaches. The simplest approach involved treatment of standard arrays with ammonium hydroxide. The second required first derivitizing slides with a photocleavable linker prior to synthesis, with oligonucleotides being ultimately recovered following a brief treatment with UV light (see **Methods**). After harvesting the oligonucleotides, we amplified the pooled material by PCR. PCR products obtained from both approaches were cloned using a Topo-TA cloning system (Invitrogen), and individual isolates were sequenced. Of the clones obtained from ammonium hydroxide-cleaved material,5 of 5 readable sequences had the correct length, correctly matched one of the sequences in the array pattern design and each sequence was unique. Of the clones obtained from photocleaved material, 4 of 5 readable sequences had the correct length and, again, each perfectly matched a unique sequence in the array pattern

design.These results suggest that the use of this highly parallel synthesis approach was feasible for producing libraries of ~60-bp fragments.

While the ability to produce complex libraries comprised of defined 60 nucleotide fragments is sufficient for some applications, others require longer oligonucleotides. For our purposes, the design of optimized shRNA libraries requires synthesis of oligonucleotides that are ~100 nucleotides in length. To test the feasibility of using microarray cloning for these purposes, we designed and printed arrays containing 96-nt sequences deposited either once per array or at variable representation (ranging from 1 to 1024 times). When we cloned PCR products derived from ammonium hydroxide-cleaved material, an average of ~63% (total of 30 in three separate cloning trials) of the sequenced clones had the correct length and correctly matched sequences printed on the array. With the arrays printed at variable oligonucleotide representation, an overwhelming majority corresponded to the sequence that was spotted 1024 times. We were unable to clone from photocleaved 96-nt material. It is possible that these longer oligonucleotides have an increased sensitivity to treatment with UV light.

Following our success in cloning long oligonucleotides, we set out to use *in situ* synthesized sequences to build shRNA expression libraries targeting nearly every identified and predicted gene in the genomes of several species including human, mouse and rat. Initially, we found that we recovered very few perfect, cloned hairpin sequences. Most of the errors were point mutations that mapped primarily to the stem regions (see Fig. 2C). This strongly suggested either that PCR selectively introduced errors during copying of structured regions or that amplification preferentially occurred on mutant templates with a lower thermal stability. Indeed, an examination of the relatively few perfect shRNAs obtained by cloning showed that we recovered only those with the lowest melting temperature (T_m). Additionally, it is well established that Taq polymerase very frequently introduces mutations if the enzyme encounters a structured region near a primer binding site [31]. Therefore, we sought to improve the amplification process by using thermostable polymerases that have proof-reading capability and that are able to effect strand displacement. Furthermore, we tested the addition of PCR enhancing agents such as DMSO or betaine. By a combination of these modifications to our procedure (see **Methods**), we were able to achieve success rates consistently ranging upward from ~25% for cloning of perfect shRNAs. These rates are similar to those that we previously obtained from cloning either individual or pools of standard synthetic oligonucleotides (see Table 1).

Thus far, we have synthesized and cloned 195,077 pooled oligonucleotides homologous to murine genes and 187,905 pooled oligonucleotides homologous to human genes. These were carried on 10 and 11 synthetic arrays, respectively (Hannon et al., unpubl.). To examine the integrity of the library shRNA populations, we sample sequenced 14,936 clones from the human library and 14,229 clones from the mouse library. Overall, we found that

20%-30% of human or mouse clones had perfect matches to the designed sequences (Table 1). Because of the difficulty in sequencing structured DNAs, these percentages probably underestimate the actual quality of these mixed libraries.

Of the 3,312 perfect shRNAs obtained from the human chip 3 sequencing, 2,635 were unique, suggesting that our approach produces relatively uniform populations. Similarly of 3,279 mouse shRNAs, 2,680 were unique. This compared favorably with results from conventionally synthesized oligos. Of 3490 sequencing runs, 936 produced perfect sequences giving a comparable overall performance for conventional and ink-jet oligonucleotides in quality. Of the 936 sequences with perfect matches, 446 were unique. It is difficult to compare the numbers of unique sequences cloned from ink-jet and conventionally synthesized populations as the pool of conventionally synthesized oligonucleotides was smaller. An examination of the T_m profile of the recovered, perfect shRNAs showed that it largely reflected the T_m profile of the total library oligonucleotide population, although there was a shift toward lower T_m for perfect clones (Fig. 2A). Similar results were obtained for conventional oligonucleotides (Fig. 2B). These results suggested that modification of the PCR procedures had somewhat, but not completely, counteracted the preference for amplification of hairpins with lower thermal stability. The difference in T_m between the perfect and expected clones represents a shift corresponding to approximately 2 additional G-C base pairs. Furthermore, an examination of the error profile of the sequences suggested that there still existed a bias for errors within the stem regions, although not nearly as pronounced as we had previously observed (Fig. 2C,D). Given that the bias toward stem region errors must arise from the amplification process, our sequencing results under-represent the overall quality of ink-jet oligonucleotides. The peaks of errors that are observed within the loop region do not correspond to any regions of known structure. All represent adenine residues, however, potentially indicating some bias in the chemical synthesis procedure.

To examine in more detail how well the cleaved and amplified oligonucleotide populations reflected the printed material, we used a standard microarray hybridization strategy. We printed and cleaved a set of oligonucleotides containing 18,723 unique 97-base oligonucleotides encoding short hairpin RNAs. Up to 3 G-C base pairs in the stem sequences of these encoded shRNAs were converted to G-U base pairs to alleviate secondary structure at the DNA level. We also printed and cleaved four oligonucleotide subsets, each containing 5,152 of the 18,723 sequences. The four subset arrays were designed such that each array overlapped the subsequent array by ~600 sequences. A T7 promoter-adapted PCR primer was used to prepare double-stranded templates for *in vitro* transcription (IVT) following cleavage and PCR. Transcription of these templates in the presence of amino allyl-UTP allowed coupling of the resulting IVT products to Cy3 and Cy5 dyes. After coupling, we hybridized dye-labeled material to a "diagnostic" microarray that contained 60mer probes of the 18,723 full-set sequences along with control sequences. To

minimize cross-hybridization, we removed the common primer binding sites from the 18,723 shRNA oligonucleotide probes on the diagnostic array.

As shown in Figure 3, we observed a single-mode distribution of hybridizing probes (high and low intensity) on the diagnostic microarray for the full-set pool and, as expected, bimodal distributions for the subset pools. After normalization for background hybridization using negative controls on the microarray, labeled IVT product from the full-set of sequences hybridized to ~99.8% of the unique sequence probes. The collective data for the four subset oligonucleotide pools revealed ~390 sequences that showed overlap in hybridization on all four arrays. This overlap was not intended in the array designs. On further inspection, it became apparent that the members of this set of sequences shared a highly conserved internal core of about 10 consecutive bases (GGGTTGGCTC) that included the conserved shRNA loop structure (see Supplementary Fig. 1). These fortuitous stretches of sequence conservation likely explain the cross hybridization observed. Of the probes on the microarray, 909 sequences contain the sequence GGGTTGGCTC from positions 27 through 36.

As a visual illustration of the coverage afforded by our library pools, we eliminated the 909 probes with the common core sequence GGGTTGGCTC and carried out a two-dimensional intensity cluster analysis of 17552 good probes (representing more than 98% of the 17898 valid probes) with the bright probes for the subset arrays. As shown in Figure 4, each cleaved subset array gave a unique signature. As expected, we observed small clusters of bright probes for each array that were also bright for intended overlapping arrays (white boxes). We used the data from the subset arrays to calculate false positive and false negative hybridization rates. We define a false positive for a subset array as a sequence determined to have significant representation in hybridization but not belonging to the 5,152 sequences actually printed on the array from which the oligonucleotide pool was obtained. We defined a false negative as a sequence that was not significantly represented in hybridization despite the fact that it was one of the intended sequences. For each subset array, the threshold for the representation significance was set such that the sum of the false positive rate and the false negative rate was minimized. The computed threshold essentially segments the bimodal probe intensity distribution into two groups, the represented sequences and the background (Fig. 3, Subset 3 data set, magenta dashed line). The same approach can be extended to the full-set array to estimate the number of sequences that are represented, in which case the representation threshold segments the full-set probes (represented) from the negative controls probes (background). With this approach, we obtained an average false positive rate of 6.15% and an average false negative rate of 1.99%. The higher, but still quite low, false positive rate likely results from a much smaller set of sequence redundancies that remain after removal of the 909 GGGTTGGCTC-containing sequences (data not shown). Thus, the true false positive rate probably approaches that of the false negative rate. Considered

together with the sample sequencing, these data suggest that pools of oligonucleotides cleaved from microarrays are extremely well represented.

Discussion

Cost-effective approaches for cloning complex libraries of predefined nucleic acid sequences are very limited. Typically, if there is no natural source of the nucleic acid, oligonucleotides must be synthesized on an individual basis for engineering into the larger library. This traditional approach is disadvantageous in several respects. First, it is costly, which limits the length and number of sequences desirable for inclusion in the library. Second, this approach is labor intensive, as each individual oligonucleotide must be manipulated for engineering into the library. Even in the case where natural sources are available, cloning and manipulation of these might not produce ideally structured populations. Our data show that microarray-based library cloning methods provide a rapid, cost-effective, flexible approach to the generation of complex, normally distributed libraries of defined oligonucleotides.

Because ink-jet microarray synthesis has been optimized for hybridization of oligonucleotides of 60 bases or fewer, we were concerned that synthesis efficiency might be severely compromised as length increased. At longer lengths, it seemed possible that truncated products might dominate the amplification steps thus diminishing the efficiency of cloning the correct length product. In light of these concerns, we were surprised to find very high fidelity clone retrieval with ink-jet synthesized oligonucleotides of up to 96 bases. Moreover, with rapidly advancing improvements to microarray synthesis technology, it is highly probable that the capability of printing much longer sequences will be feasible in the near future.

We were also surprised to note the lack of bias in the amplification of complex populations. Complex pools of PCR templates with non-degenerate primer binding sites are rarely used for amplification. Therefore, we were concerned that specific sequences might dominate our amplified and cloned material. Initially we did see biases against recovery of templates with the most stable secondary structures. However, with modifications of our amplification procedures, these were effectively eliminated as measured both by sample sequencing and array hybridization.

In mammals, shRNAs have proven an effective method of triggering an RNAi response, resulting in gene silencing [32-38]. We are presently applying ink-jet synthesis to the production of defined, informatically optimized, genome-wide shRNA libraries (Hannon et al., unpublished). However, this technology can be applied to the construction of oligonucleotide populations for other purposes. For example, this method would be ideal for generating libraries for antibody diversity studies, phage display, combinatorial peptide sequence generation, DNA binding site selection, promoter region analysis, and restriction enzyme site analysis.

As demonstrated above, *in situ* ink-jet synthesis of oligonucleotide populations can be used to generate cloned libraries. Alternatively, either the primary synthesis product or an amplified population can be used without cloning. Such an application is exemplified by the use of the cleaved and amplified oligonucleotide population as a template for RNA synthesis by T7 polymerase. Although the resulting products could be used for many purposes, one intriguing possibility is the use of such populations to validate microarray probe sets for mRNA expression analysis of genome-wide chromatin immunoprecipitation. Notably, given the apparently even distribution of sequences within the synthesized population, complex libraries of defined oligonucleotides would serve ideally as a common reference population for ratio microarray experiments.

Given the accuracy and flexibility of ink-jet oligonucleotide synthesis, it seems likely that this approach will become a preferred method for constructing diverse library-based tools for functional genomic studies. We have demonstrated that this technology allows the generation of very large numbers of defined oligonucleotides at an almost incidental cost. Such flexibility will greatly enhance the speed with which advances in dissecting gene function can be made and further refined, thereby accelerating achievement of the goals of the post-genomic era.

Methods

Oligonucleotide Design and Microarray Synthesis

Sequences to be included in a library were designed such that each was flanked by 5' and 3' common 14- to 18-base PCR primer recognition sites. Oligonucleotide microarrays were printed at Agilent Technologies or synthesized at Rosetta using ink-jet technology as described previously [27]. Prior to harvesting the oligonucleotides, quality control testing was performed using a functional hybridization of representative arrays that were produced on the same manufactured glass substrates.

Oligonucleotide Cleavage with a Photocleavable Spacer

Photocleavable spacer phosphoramidite (Glen Research, VA) monomers were synthesized on a silanized 3" x 3"x 0.004" glass wafer with hydroxyl functionality. Silanization of glass surfaces for oligonucleotide applications have been described [28,30] and silanes with various functionality are commercially available (Gelest, PA). All reaction steps and reagent preparations were performed under nitrogen in a PLAS-LABS, 830-ABC glove box (PLAS-LABS, MI). Anhydrous acetonitrile (1 mL; Fisher Scientific, NH) was added via syringe injection to 100 μmoles of freeze-dried photocleavable spacer phosphoramidite to yield a 0.1 M solution. Anhydrous acetonitrile (62 mL) was then added to 2 g

of freeze-dried 5-ethylthiol-1H-tetrazole (Glen Research, VA) to yield a 0.25 M solution for phosphoramidite activation. The solutions were vortexed briefly and allowed to equilibrate at room temperature for 30 min. The tetrazole solution (1 mL) was transferred by syringe to the photocleavable spacer solution and the mixture vortexed for 10 sec. Two silanized wafers were placed 'reactive side up' and 2 mL of the active photocleavable spacer/tetrazole solution was added to the surface of the first wafer. The second wafer was placed sandwich-like on the first, allowing the fluid to distribute uniformly between the surfaces. The wafers were incubated at room temperature for 2 min, separated, placed in a Teflon™ rack and immersed in a bath of acetonitrile. The rack was agitated in the bath for 2 min to ensure complete rinsing of excess photocleavable spacer and dried by centrifugation. Formation of the stable pentavalent phosphodiester and removal of the dimethoxytrityl protecting group were carried out per standard oligonucleotide synthesis procedures [27,29]. Synthesis of oligonucleotides on photocleavable spacer-functionalized substrates was performed as described above.

For arrays synthesized with a photocleavable spacer, the oligonucleotides were cleaved in 1 mL of 25 mM Tris-buffer solution (pH 7.4) under UV irradiation (300-nm wavelength) for 30 min. The solution was transferred to a 1.5-mL microcentrifuge tube and speed vacuumed at low heat overnight.

Oligonucleotide Cleavage Using Ammonium Hydroxide

To cleave oligonucleotides synthesized without a photocleavable spacer, the microarrays were treated for 2 hrs with 2-3 mL of 35% NH_4OH solution (Fisher Scientific) at room temperature. The solution was transferred to 1.5-mL microcentrifuge tubes and speed vacuum dried at medium heat (~55 °C) overnight.

PCR Amplification of Cleaved Oligonucleotides.

Dried material containing oligonucleotides cleaved from each microarray was resuspended in 250 μl of RNase/DNase-free H_2O. For PCR template, a range of volumes (0.1-5.0 μl) was tested to determine the amount that gave the best yield with the lowest incidence of non-specific product. We carried out PCR amplification of the initial 59 and 96 nt test sequences in 50 μl reactions containing 1x PCR buffer minus Mg (Invitrogen), 9% sucrose, 1.5 mM $MgCl_2$, 1 ng/μl forward and reverse primers, 125 μM dNTPs, and 0.05 U/μl Taq polymerase. Thermocycler conditions depended on the length of the oligonucleotides and the melting temperatures of the forward and reverse primers. In general, 30 cycles of 94°C denaturing for 30 sec, annealing at the appropriate temperature for 30 sec, and extension at 72°C for 90 sec worked well. If the PCR products were to be cloned using a TA cloning system such as the Topo/TA cloning system (Invitrogen), we used Taq polymerase and followed the 30-cycle PCR with a 10 min extension at 72°C. For the cloning of shRNA

libraries, the use of Vent polymerase or Pfx polymerase in the presence of DMSO and/or betaine reduced the incidence of nucleotide misincorporation during the PCR. We optimized conditions separately for each primer set used. In some cases, PCR products were cleaned up by gel purification using the QIAquick Gel Extraction protocol (QIAgen). In other cases, the PCR products were simply cleaned up following a QIAquick PCR purification protocol (QIAgen).

Reverse Transcription/In Vitro Transcription (RT/IVT) and Microarray Hybridization

To prepare templates for T7 *in vitro* transcription, we pooled PCR material from two individual reactions. Unincorporated nucleotides and polymerase were removed from the pooled PCR products by QIAquick PCR purification (QIAgen) with elution in 50 μl of RNAse/DNase-free water. Eluates were speed-vacuum dried to concentrate two-fold and 7.25 μl was used as template in a T7 RNA polymerization reaction using a modified Megashortscript protocol (Ambion). In lieu of 2 μl of 75 mM UTP, we used 2.25 μl of 50 mM amino allyl UTP (aa-UTP; Ambion) plus 0.5 μl of the 75 mM UTP provided with the kit. The reactions were carried out at 37°C overnight. Then, 1 μl of DNase was added for 15 min at room temperature. Next, the samples were phenol/chloroform/isoamyl alcohol extracted and ethanol precipitated. Final resuspension was in 40 μl of water.

Amino allyl-UTP incorporated cRNA was aliquoted into two, 96-well plates (5 □g per reaction well). One plate for Cy3 NHS-esther coupling and one for Cy5 NHS-esther coupling were prepared (dyes were obtained from Amersham Biosciences, NJ). Samples were reacted with the dyes and mixed for performance of two color ratio experiments and subsequently purified using BIO-RAD Micro Bio-Spin columns P-30 Tris (Bio-Rad Laboratories, CA). Purified dye-labeled samples were then hybridized for 24 hrs to the detection microarray, washed, scanned on an Agilent Scanner and analyzed. Rosetta standard coupling and hybridization processes were employed as previously described [27].

Acknowledgements

G.J.H. is supported by an Innovator Award from the U.S. Army Breast Cancer Research Program. This work was also supported by a grant from the NIH (GJH). We thank Hongyue Dai for suggestions regarding microarray analysis of the library population and the Rosetta Gene Expression Laboratory for microarray RNA processing and hybridizations.

Table 1 *Characterization of cloned shRNAs.*

Material Source	Sequencing runs	Perfect match	1 mismatch	2 mismatches	>2 mismatches	Percent perfect
Ink-jet human chip 1	2287	701	668	265	633	30.6%
Ink-jet human chip 3	12749	3312	3484	1715	4238	26%
Ink-jet mouse chip 4	14229	3279	4319	2213	4418	23%
Conventional synthesis (commercial vendor)	3304	936	457	310	1601	28.3%

Figure Legends

Figure 1.

Cloning strategy using in situ oligonucleotide synthesis. To create a pool of sequences for library cloning, oligonucleotides are printed on a microarray substrate, cleaved by strong base treatment or ultraviolet light and amplified by PCR. The amplified products are treated with restriction enzymes or use directly for ligation as a pool into the vector of choice.

Figure 2.

Characterization of shRNA cloning from in situ oligonucleotides. **A,B.** T_m profiles of sequenced clones that perfectly matched the expected sequences (green) are compared with the T_m profile of the entire library (red) for ink-jet (A) or conventionally synthesized (B) oligonucleotides. The entire population of library oligonucleotides in A was 195077 sequences compared with 15519 correct clones; in B the entire library was 1995 sequences compared with 1380 correct clones. T_ms were calculated according to Turner [39]. **C,D.** The nucleotide positions of errors in incorrect sequences were mapped in the shRNA template for ink-jet (C) or conventionally synthesized (D) oligonucleotides. The stem and loop regions of the template are indicated diagrammatically. (Red) Traces from human library oligonucleotides; (green) traces from mouse library oligonucleotides. In C, 37020 human and 9829 mouse library traces were analyzed; in D, 2772 human library traces were analyzed.

Figure 3.

Histograms of the average intensity of the 18,723 probes when hybridized to IVT products derived from the pool of full-set of sequences (left) and one

representative subset of 5152 sequences (right). Subsets arrays 1, 2, and 4 showed similar bimodal distributions.

Figure 4

The subset sequences give unique signatures of bright intensity probes and show the expected overlap. The heat map shows the results of two-dimensional clustering of logarithmic intensities of 17552 good probes, representing >98% of the 17898 valid probes (excluding 909 total GGGTTGGCTC-containing sequences) on the full set and subset cloning array samples. (Pink) Bright intensity probes; (black) dim intensity probes. (White boxes) Probes with expected overlap among the subset arrays. Note that the probe intensity from each array is normalized by its computed threshold for representation so that a sequence is considered represented when its logarithmic intensity is >0.

Supplementary Figure 1.

Sequence traces of 390 oligonucleotides with a highly conserved core region of about 10 consecutive bases (GGGTTGGCTC) that includes the shRNA loop structure. This conservation is thought to underlie the cross hybridization observed with material from the subset array pools.

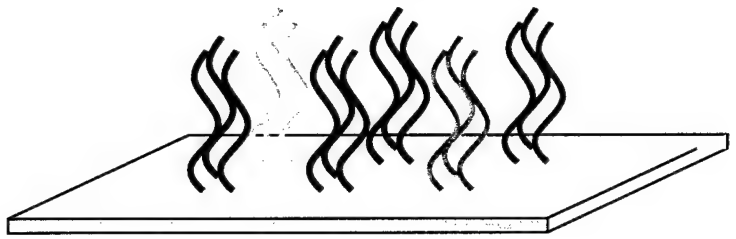
REFERENCES

1. Hudson, J. D. et al. A proinflammatory cytokine inhibits p53 tumor suppressor activity. *J Exp Med* **190**, 1375-82 (1999).
2. Brummelkamp, T. R. et al. TBX-3, the gene mutated in Ulnar-Mammary Syndrome, is a negative regulator of p19ARF and inhibits senescence. *J Biol Chem* **277**, 6567-72 (2002).
3. Maestro, R. et al. Twist is a potential oncogene that inhibits apoptosis. *Genes Dev* **13**, 2207-17 (1999).
4. Raveh, T., Berissi, H., Eisenstein, M., Spivak, T. & Kimchi, A. A functional genetic screen identifies regions at the C-terminal tail and death-domain of death-associated protein kinase that are critical for its proapoptotic activity. *Proc Natl Acad Sci U S A* **97**, 1572-7 (2000).
5. Fletcher, B. S., Dragstedt, C., Notterpek, L. & Nolan, G. P. Functional cloning of SPIN-2, a nuclear anti-apoptotic protein with roles in cell cycle progression. *Leukemia* **16**, 1507-18 (2002).
6. Rayner, J. R. & Gonda, T. J. A simple and efficient procedure for generating stable expression libraries by cDNA cloning in a retroviral vector. *Mol Cell Biol* **14**, 880-7 (1994).
7. Kitamura, T. et al. Efficient screening of retroviral cDNA expression libraries. *Proc Natl Acad Sci U S A* **92**, 9146-50 (1995).
8. Kojima, T. & Kitamura, T. A signal sequence trap based on a constitutively active cytokine receptor. *Nat Biotechnol* **17**, 487-90 (1999).
9. Golovkina, T. V. et al. A novel membrane protein is a mouse mammary tumor virus receptor. *J Virol* **72**, 3066-71 (1998).
10. Battini, J. L., Rasko, J. E. & Miller, A. D. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: possible role in G protein-coupled signal transduction. *Proc Natl Acad Sci U S A* **96**, 1385-90 (1999).
11. Gallagher, W. M., Cairney, M., Schott, B., Roninson, I. B. & Brown, R. Identification of p53 genetic suppressor elements which confer resistance to cisplatin. *Oncogene* **14**, 185-93 (1997).
12. Garkavtsev, I., Kazarov, A., Gudkov, A. & Riabowol, K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. *Nat Genet* **14**, 415-20 (1996).
13. Hannon, G. J. RNA interference. *Nature* **418**, 244-51. (2002).
14. Fraser, A. G. et al. Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature* **408**, 325-30. (2000).
15. Gonczy, P. et al. Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. *Nature* **408**, 331-6. (2000).
16. Lee, S. S. et al. A systematic RNAi screen identifies a critical role for mitochondria in C. elegans longevity. *Nat Genet* **33**, 40-8 (2003).
17. Lum, L. et al. Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. *Science* **299**, 2039-45 (2003).

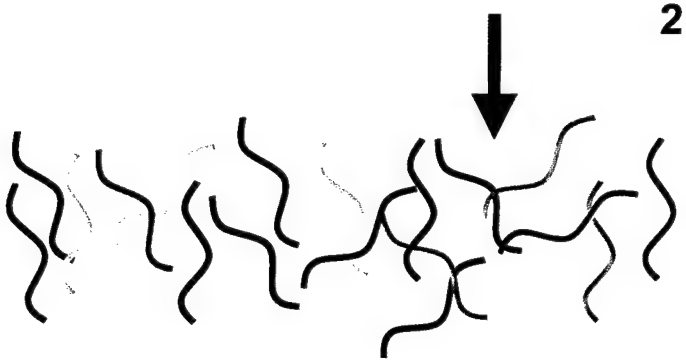
18. Paddison, P. J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427-31 (2004).
19. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-7 (2004).
20. Sen, G., Wehrman, T. S., Myers, J. W. & Blau, H. M. Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat Genet* **36**, 183-9 (2004).
21. Shirane, D. et al. Enzymatic production of RNAi libraries from cDNAs. *Nat Genet* **36**, 190-6 (2004).
22. Luo, B., Heard, A. D. & Lodish, H. F. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc Natl Acad Sci U S A* **101**, 5494-9 (2004).
23. Hsieh, A. C. et al. A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res* **32**, 893-901 (2004).
24. Schwarz, D. S. et al. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).
25. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-16 (2003).
26. Jackson, A. L. et al. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**, 635-7 (2003).
27. Hughes, T. R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**, 342-7 (2001).
28. Halliwell, C. & Cass, A. E. A factorial analysis of silanization conditions for the immobilization of oligonucleotides on glass surfaces. *Anal. Chem.* **73**, 2476-2483 (2001).
29. Brown, D. M. A brief history of oligonucleotide synthesis. *Methods Mol Biol* **20**, 1-17 (1993).
30. Bourdeiu, L., Silberzan, P. & Chatenay, D. Langmuir-Blodgett films: From micron to angstrom. *Physical Review Letters* **7**, 2029-2032 (1991).
31. Loewen, P. C. & Switala, J. Template secondary structure can increase the error frequency of the DNA polymerase from Thermus aquaticus. *Gene* **164**, 59-63 (1995).
32. Brummelkamp, T. R., Bernards, R. & Agami, R. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **21**, 21 (2002).
33. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev* **16**, 948-58. (2002).
34. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nat Biotechnol* **20**, 505-8. (2002).
35. Lee, N. S. et al. Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol* **20**, 500-5. (2002).
36. Sui, G. et al. A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc Natl Acad Sci U S A* **99**, 5515-20. (2002).

37. Kawasaki, H. & Taira, K. Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Res* **31**, 700-7. (2003).
38. Miyagishi, M. & Taira, K. U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat Biotechnol* **20**, 497-500. (2002).
39. Turner, D. H. Thermodynamics of base pairing. *Curr Opin Struct Biol* **6**, 299-304 (1996).

1. Synthesize array with sequences to clone



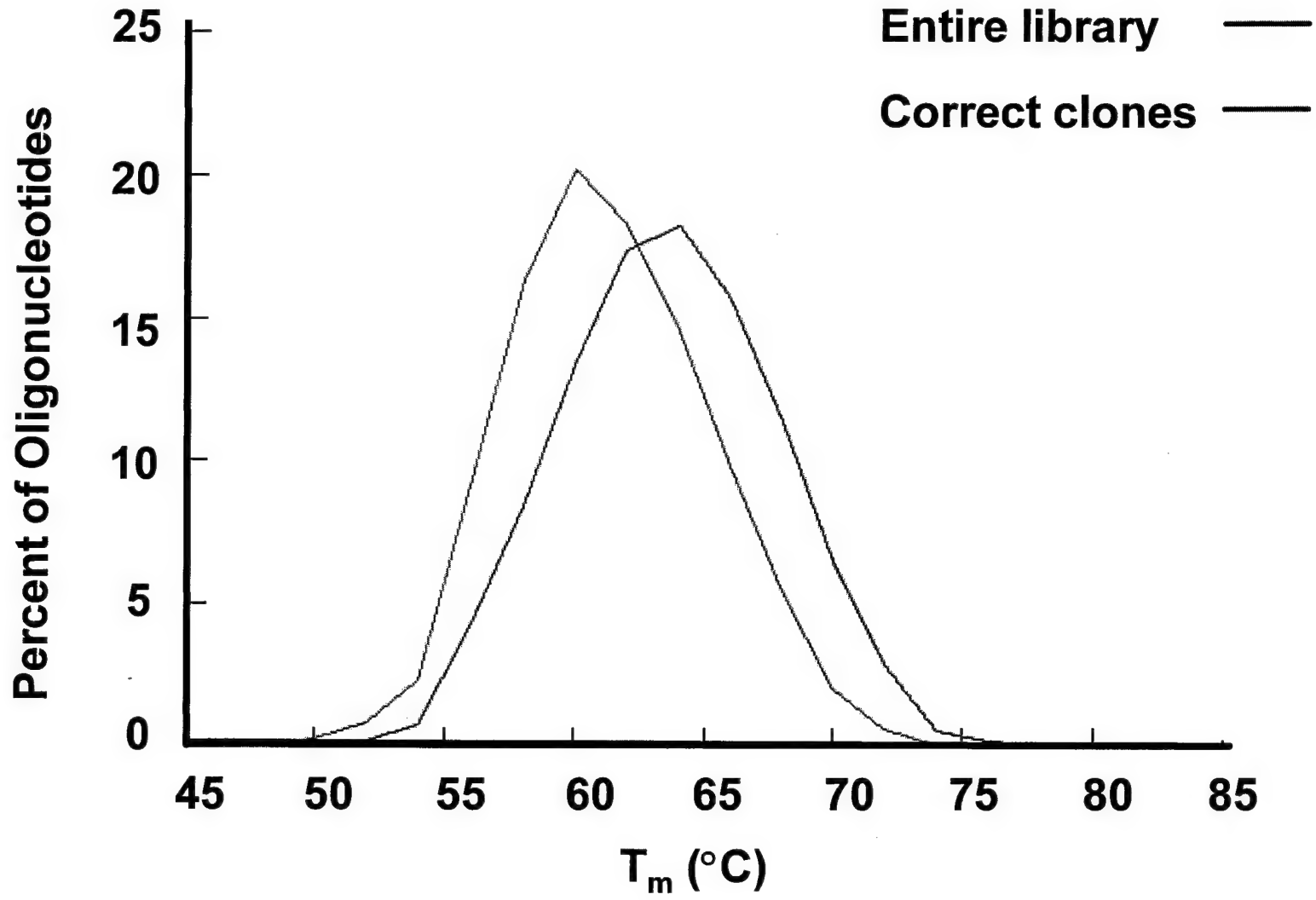
2. Treat array to harvest oligos



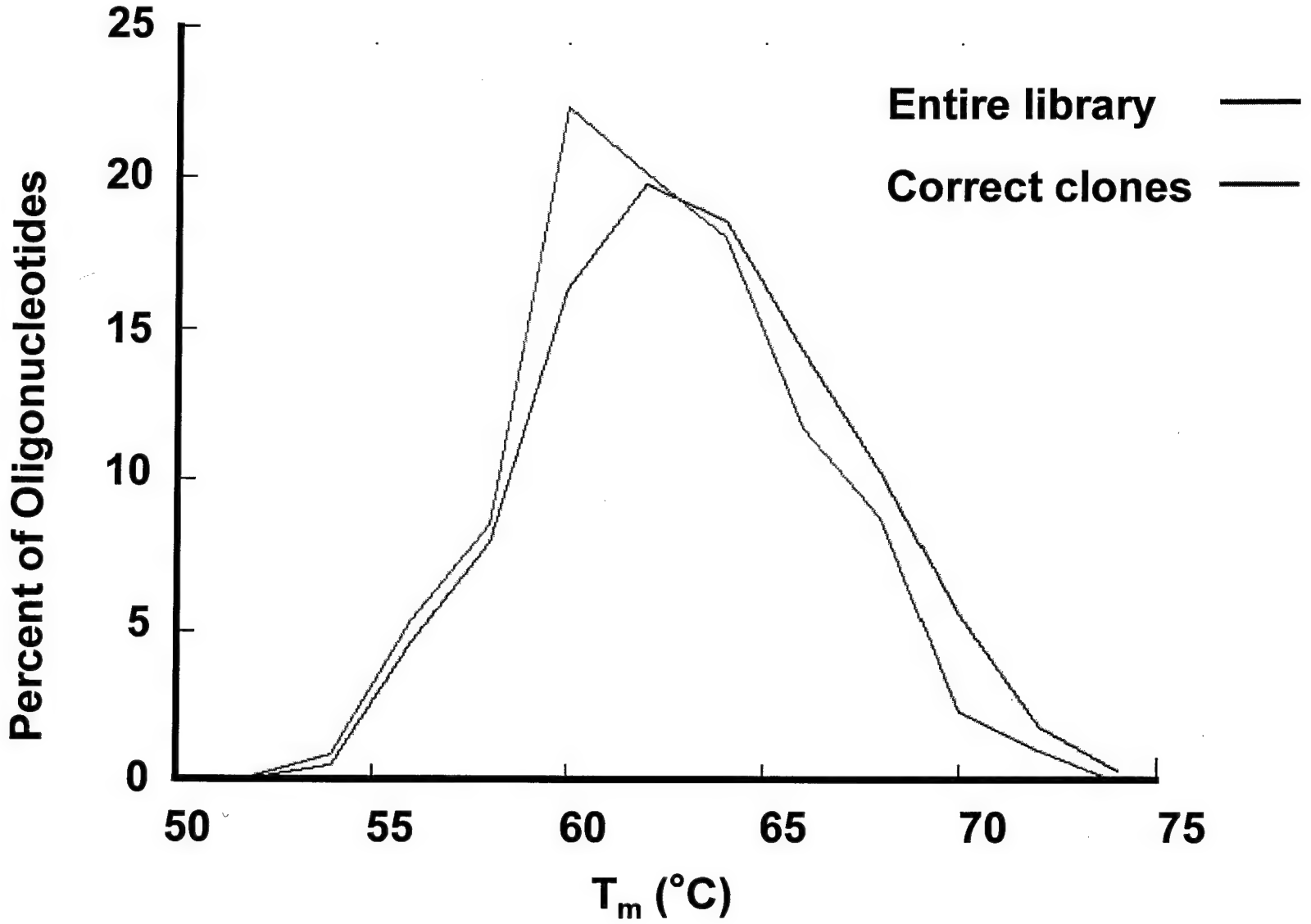
3. PCR amplify to generate dsDNA

4. Clone PCR products to generate library

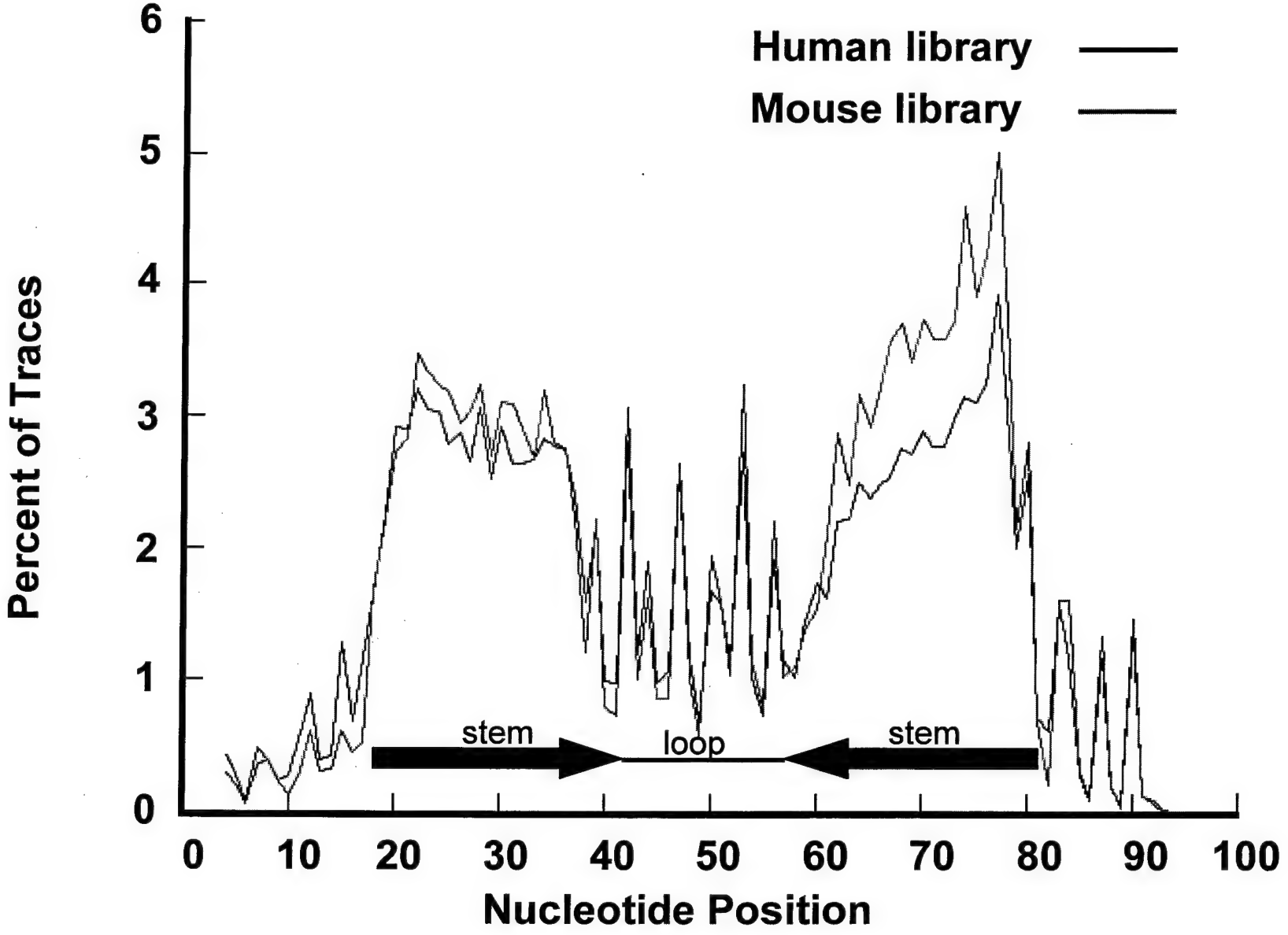
Cleary et al., Fig 2A



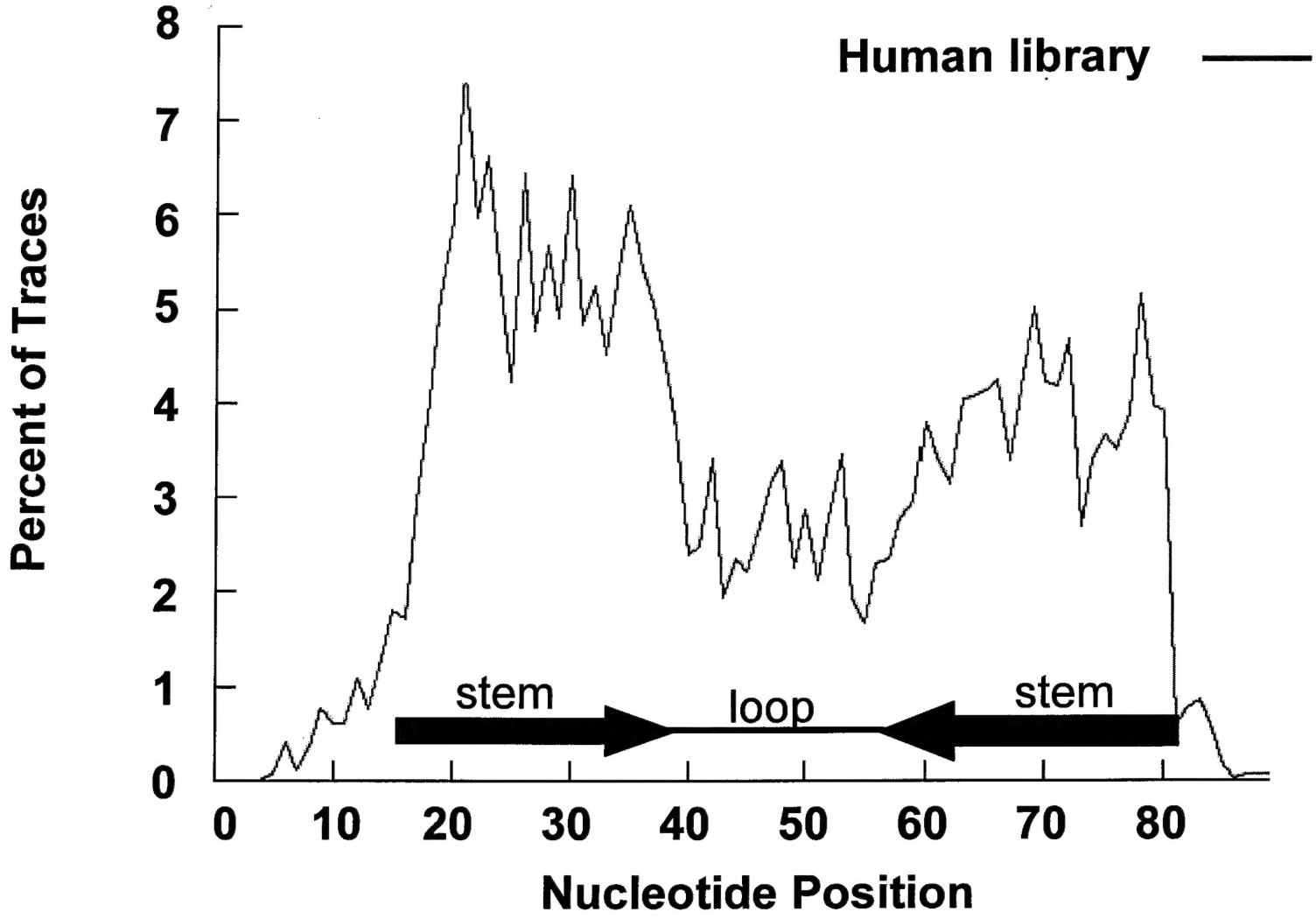
Cleary et al., Fig 2B

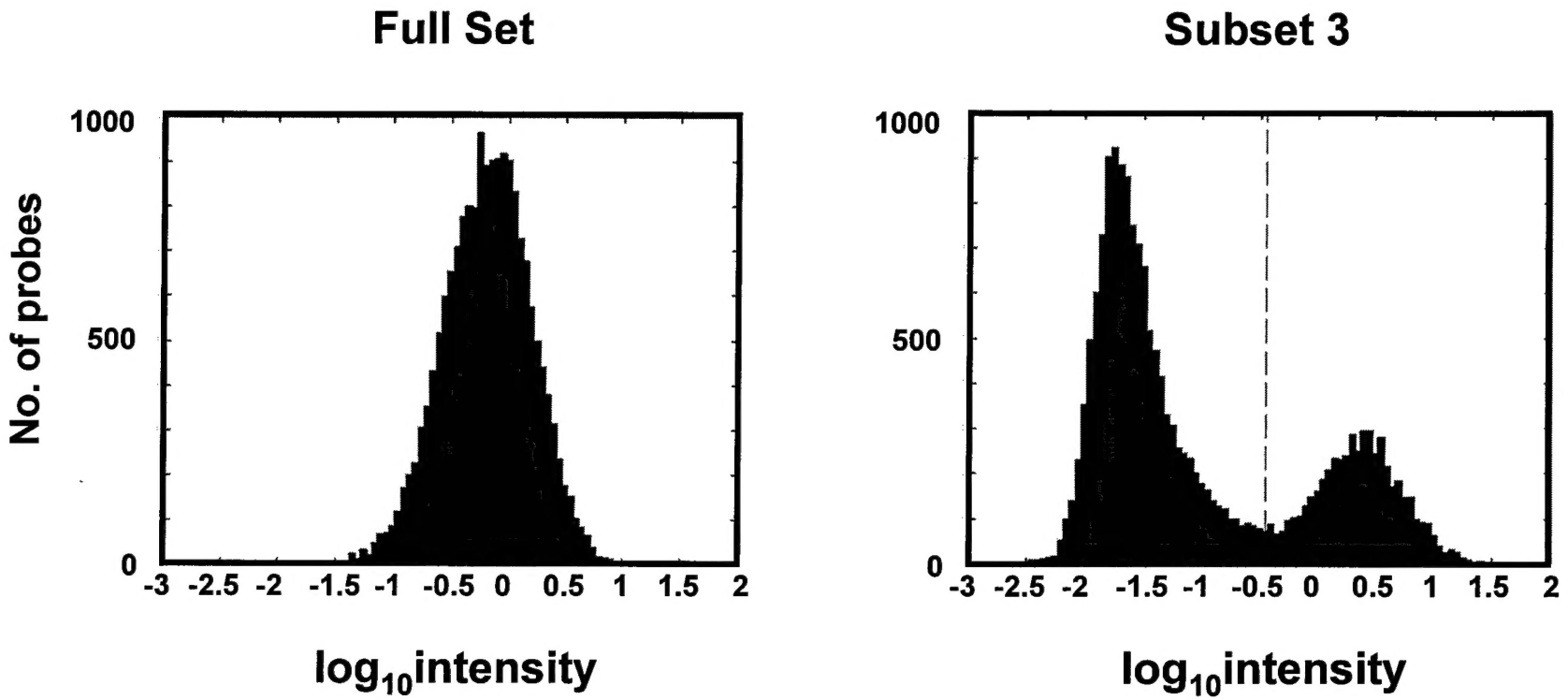


Cleary et al., Fig 2C

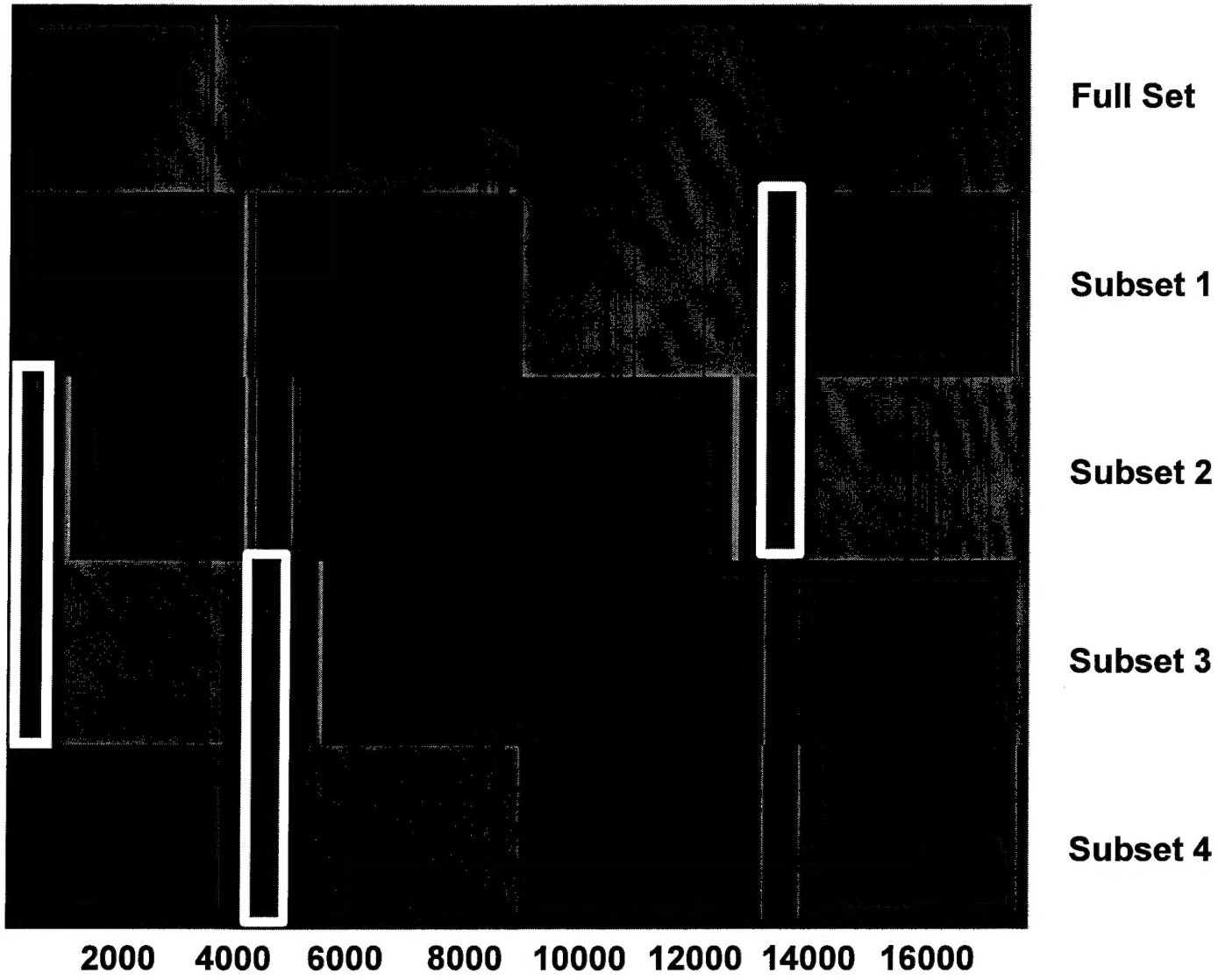


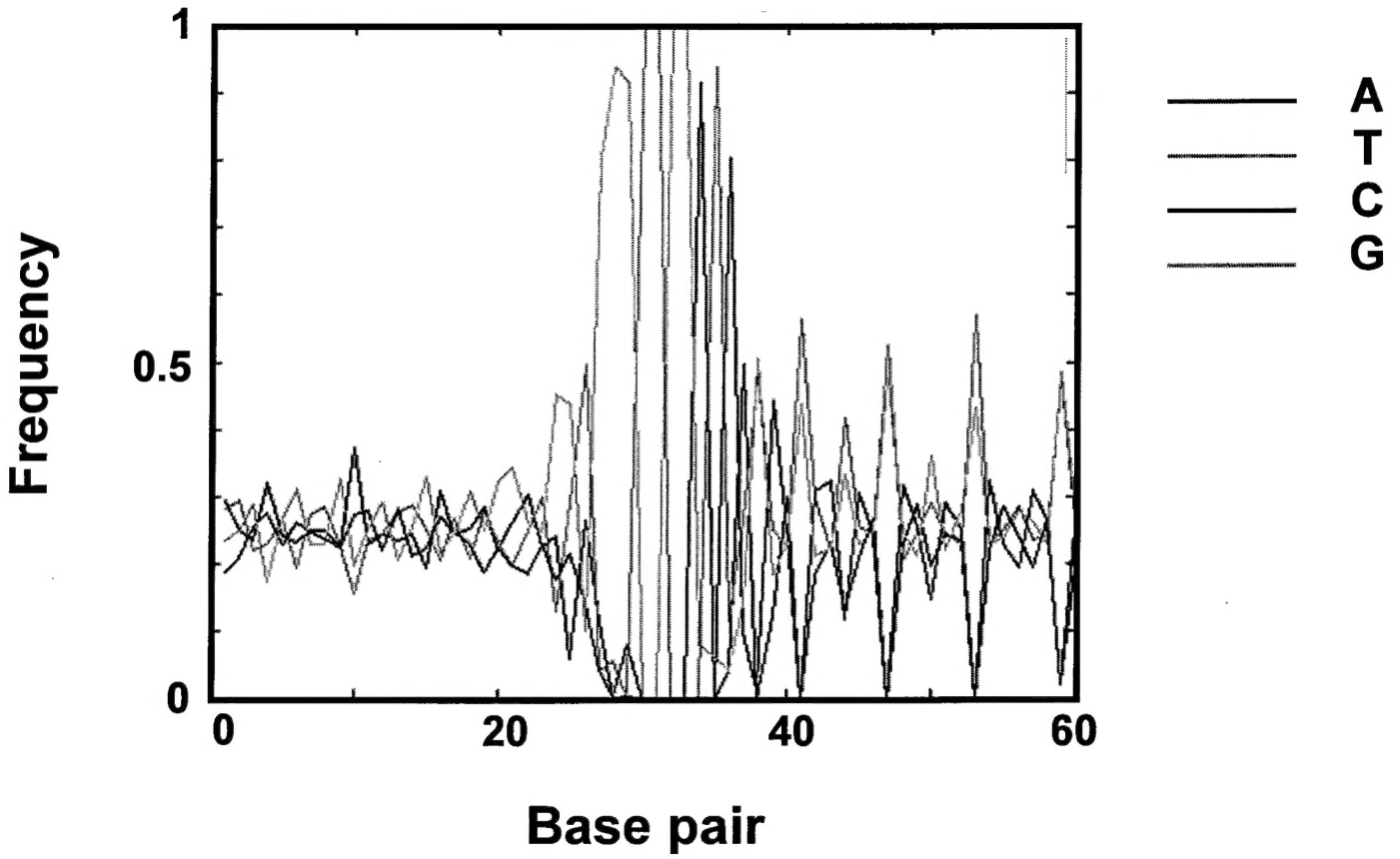
Cleary et al., Fig 2D




Full Set
Subset 3
No. of probes
1000
500
0
-3 -2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2
log10intensity

Cleary et al., Fig 4




Frequency
1
0.5
0
0
20
40
60
Base pair
A
T
C
G